

Upper tail dependence in rainfall extremes: would we know it if we saw it?

Francesco Serinaldi · András Bárdossy ·
Chris G. Kilsby

Published online: 11 September 2014

© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract The simultaneous occurrence of extreme events, such as simultaneous storms and floods at different locations, has a serious impact on risk assessment and mitigation strategies. The joint occurrence of extreme events can be measured by the so-called upper tail dependence (UTD) coefficient λ_U . In this study, we reconsider the properties of the most popular λ_U estimators and show that their strong bias and uncertainty make most of the empirical results reported in the hydrological literature questionable. In order to overcome the limits of λ_U analysis, we test several alternative tools such as a pool of formal statistical tests devised for recognizing upper tail independence and graphical diagnostics based on binary correlation and binary entropy. The reliability of all the methods is preliminarily checked by Monte Carlo experiments. Statistical tests and graphical diagnostics are therefore applied to three different rainfall data sets that allow us to explore the properties of the spatial dependence structure of rainfall extremes over a wide range of spatio-temporal scales ranging from 30 min and 1 km to 30 days and ≈ 3000 km. Results highlight that (1) classical estimators provide non zero tail dependence even for cases where it should be zero; (2) formal tests and binary correlation highlight that the pairwise spatial dependence structure can be weaker than Gaussian, thus excluding UTD

calculated in a pairwise manner; (3) the binary entropy computed on triples of locations shows that the pairwise UTD is not enough to explain the spatial dependence structure of extreme rainfall, whose complexity becomes evident only after resorting to higher order correlation measures. The results concerning the bias and uncertainty of λ_U estimators are fully general and suggest avoiding their use especially for the short time series usually available in hydrology.

Keywords Upper tail dependence · Extreme events · Binary correlation · Binary entropy · Rainfall · Peak over threshold analysis · Collective spatial risk

1 Introduction

The fast growth of multivariate frequency analysis (thanks partly to the introduction of apparently more manageable statistical tools such as copulas) has led to an extensive application of multivariate models to a variety of hydrological analyses going from the study of the relationships between the characteristics of objects such as drought events and hydrographs (e.g., Serinaldi et al. 2009; Volpi and Fiori 2012) to the study of the occurrence of extreme events at multiple sites (e.g., Ghizzoni et al. 2010) to spatial interpolation and simulation problems (e.g., Bárdossy 2006; Bárdossy and Li 2008; Bárdossy and Pegram 2013). This intense activity resulted in a large body of literature that was almost unavoidably focused on showing the potential application rather than on the actual nature of the variables at hand, the possible shortcomings of the methods used, and the reliability of multivariate methods applied to the usually very short hydrological time series.

In this context, Serinaldi and Kilsby (2013) and Serinaldi (2013) attempted to re-focus the attention on some basic concepts such as a careful examination of the nature of the

F. Serinaldi (✉) · A. Bárdossy · C. G. Kilsby
School of Civil Engineering and Geosciences, Newcastle
University, Newcastle Upon Tyne NE1 7RU, UK
e-mail: francesco.serinaldi@newcastle.ac.uk

F. Serinaldi · C. G. Kilsby
Willis Research Network, 51 Lime St., London EC3M 7DQ, UK

A. Bárdossy
Institute of Hydraulic Engineering, University of Stuttgart,
Pfaffenwaldring 61, 70569 Stuttgart, Germany

relationships between some hydrological variables and on the uncertainty and reliability of multivariate inference performed on typically short time series. The aim was to transfer concepts well-known in univariate analyses to the multivariate framework, where they seem to be overlooked or forgot (see also Serinaldi 2014).

This study attempts to further extend this inquiry focusing on the UTD coefficient λ_U , which is another concept related to copula theory whose estimation is nowadays a standard step in multivariate frequency analyses reported in hydrological literature. λ_U is an index that quantifies the limit probability that a variable X_1 exceeds a given t quantile x_1 given that another variable X_2 exceeds its t quantile x_2 as $t \rightarrow 1^-$. More formally, given the conditional probability $\mathbb{P}[F_{X_1}(x_1) > t | F_{X_2}(x_2) > t]$ and introducing the copula formalism $F_{X_1 X_2}(x_1, x_2) = C(F_{X_1}(x_1), F_{X_2}(x_2)) = C(u_1, u_2)$, where F_{X_i} , $i = 1, 2$, denotes the univariate marginal distribution of the generic variable x_i and $U_i := F_{X_i}$, λ_U is defined as (e.g., Frahm et al. 2005)

$$\lambda_U := \lim_{t \rightarrow 1^-} \mathbb{P}[F_{X_1}(x_1) > t | F_{X_2}(x_2) > t] = \lim_{t \rightarrow 1^-} \frac{1 - 2t + C(t, t)}{1 - t}. \quad (1)$$

Since λ_U gives a measure of the tendency of observing simultaneous extreme events exceeding a given quantile threshold, its evaluation is potentially of great interest to better understand if and how extreme events cluster together and to choose a suitable multivariate model. The use of λ_U in hydrological analyses can be dated back to the works of Poulin et al. (2007) and Serinaldi (2008) concerning bivariate frequency analyses of the annual maximum flows and the corresponding flow hydrograph volumes, and the pairwise analysis of rainfall data at multiple locations, respectively. As for the introduction of copulas in hydrology, the theoretical apparatus used in those works was essentially borrowed from econometric literature, namely, Schmidt (2003), Frahm et al. (2005) and Schmidt and Stadtmüller (2006). These works provide quite extensive simulation studies devised to assess the reliability of a set of parametric and nonparametric λ_U estimators as well as several conclusive warnings about their use in practical analyses. Poulin et al. (2007) made an ad hoc Monte Carlo experiment to choose the most appropriate estimator for a specific case study, retaining the Coles–Heffernan–Tawn λ_U^{CHT} estimator (Coles et al. 1999) and Capéraà–Fougères–Genest λ_U^{CFG} estimator (Capéraà et al. 1997). Poulin et al. (2007) also stressed the caveats previously reported by Schmidt (2003) and Frahm et al. (2005). Serinaldi (2008) exploited the relationship between λ_U and Kendall correlation coefficient τ_K to build a diagnostic plot useful for the model selection. In order to understand the reliability of that graphical tool, Serinaldi (2008) made a limited

simulation experiment that confirmed the bias of λ_U^{CFG} and the uncertainty characterizing the Schmidt–Stadtmüller λ_U^{SS} estimator (Schmidt and Stadtmüller 2006) (see Fig. 1 in Serinaldi 2008). However, Serinaldi (2008) did not perform an extensive assessment and reported only partially the caveats stated by Schmidt (2003) and Frahm et al. (2005). Unfortunately, the subsequent hydrological literature relied on the first papers dealing with λ_U in the same field, progressively overlooking the original theoretical works. This habit led to a rather blind use of the λ_U^{CFG} and other λ_U estimators whose application is justified by sentences such as “The [λ_U^{CFG}] estimator is based on the assumption that the empirical copula can be approximated by an extreme value copula. It also works well when this hypothesis is not fulfilled, except in the case that the real UTD is null” (Requena et al. 2013) or “Though the estimator assumes that the underlying copula can be approximated by an extreme-value copula, studies have shown that the estimator performs well even if the copula does not belong to extreme value classes (Frahm et al. 2005)” (Villarini et al. 2008; Serinaldi et al. 2009; Janga Reddy and Singh 2014). Serinaldi and Kilsby (2014a) showed how rainfall fields simulated by meta-Gaussian spatial dependence structure exhibit λ_U estimates comparable with those of the observed rainfall fields, thus suggesting caution about the recommendation of AghaKouchak et al. (2013) of using λ_U estimates to assist in planning and policy making as well as validating numerical models.

In order to shed light on such a matter, we recall that the recommendations provided by Schmidt (2003), Schmidt and Stadtmüller (2006), Poulin et al. (2007), and Frahm et al. (2005) are based on simulation experiments in which a unique value of the overall correlation is used, namely $\tau_K = 1/3$ in Schmidt (2003) and Frahm et al. (2005), Pearson correlation $\rho_P = 0.25$ in Schmidt and Stadtmüller (2006) and $\tau_K = 0.51$ in Poulin et al. (2007). Based on these simulation settings, these studies conclude that:

- (1) Among the nonparametric λ_U estimators, the λ_U^{CFG} estimator does well, although Frahm et al. (2005) advised caution regarding the sometimes low variance relative to bias. Further, λ_U^{CFG} shows a weak performance in the case of tail independence.
- (2) Among the nonparametric λ_U estimators, λ_U^{CFG} shows the best performance, whereas for (semi-)parametric estimations a specific copula (such as the t-copula) is recommended.
- (3) The nonparametric estimators are too sensitive in case of small sample sizes. Thus, under these circumstances, a parametric λ_U estimation might be favorable in order to increase the stability of the estimation although the model error could be large.

Tests for tail dependence are absolutely mandatory for every λ_U estimation because (1) samples exist which seem to be tail dependent but they are realizations of a tail independent distribution (such as a finite mixture of multivariate Gaussian distributions) and vice versa; (2) the use of misspecified parametric marginals instead of empirical margins may lead to wrong interpretations of the dependence structure; (3) the λ_U estimators can be insensitive to upper tail independence, thus indicating upper tail dependence even if no exists.

Starting from the above remarks and the need for further investigation stressed by AghaKouchak et al. (2010), in this study we investigate the actual magnitude of bias and uncertainty of two λ_U estimators highlighting the strong dependence of the estimates on the overall correlation. Therefore, following the last point of the above list, we consider five different formal tests devised for checking the hypothesis of upper tail independence, assessing their reliability by extensive Monte Carlo experiments. We also introduce two further diagnostics based on pairwise binary correlation and binary entropy on triples (Bárdossy and Pegram 2009). Finally, we apply all the diagnostics on three rainfall data sets that allow us to obtain a comprehensive picture of the joint behavior of rainfall extremes over a wide range of spatial and temporal scales.

The paper is organized as follows. Tests and diagnostics are introduced in Sect. 2 along with simulations from theoretical models. The data sets are presented in Sect. 3, whereas empirical results are reported in Sect. 4. Discussion and conclusions are summarized in Sects. 5 and 6, respectively.

2 Upper tail diagnostics

2.1 λ_U estimators

In hydrological studies, four different estimators have been commonly considered (Poulin et al. 2007; Villarini et al. 2008; Serinaldi et al. 2009; AghaKouchak et al. 2010, 2013; Janga Reddy and Singh 2014, among others): λ_U^{SS} (Schmidt and Stadtmüller 2006), the “secant-based” estimator λ_U^{SEC} (Joe et al. 1992), λ_U^{CFG} (Capéraà et al. 1997; Frahm et al. 2005), and λ_U^{CHT} (Coles et al. 1999; Frahm et al. 2005). As was summarized by Villarini et al. (2008), the first method is unbiased but can show high variance; the second estimator can be interpreted as the slope of the secant along the copula diagonal [close to the point (1, 1)], and therefore it can misspecify the value of λ_U when data are not accumulated along the diagonal; the third one

assumes that the empirical copula function approximates an extreme value (EV) copula but the estimator could be biased and show very low variance; while the fourth one is the nonparametric counterpart of the χ estimator proposed by Coles et al. (1999). In this study, we focus on λ_U^{SS} because it does not require any distributional assumption and is unbiased (for truly tail dependent models) and on λ_U^{CFG} because of its popularity related to its closed form formula. In more detail, λ_U^{CFG} is defined as

$$\lambda_U^{CFG} = 2 - 2 \left[\frac{1}{N} \sum_{i=1}^N \log \left(\frac{\sqrt{\log\left(\frac{1}{u_{1,i}}\right) \log\left(\frac{1}{u_{2,i}}\right)}}{\log\left(\frac{1}{(\max(u_{1,i}, u_{2,i}))^2}\right)} \right) \right], \tag{2}$$

where N is the sample size and $u_{1,i} := F_{X_1}(x_{1,i})$ and $u_{2,i} := F_{X_2}(x_{2,i})$, for $i = 1, \dots, N$, whereas λ_U^{SS} is

$$\lambda_U^{SS} = \frac{N}{\kappa} \bar{C}_N\left(\frac{\kappa}{N}, \frac{\kappa}{N}\right) \approx \frac{1}{\kappa} \sum_{i=1}^N \mathbb{1}_{\{R_{1,i} > N-\kappa, R_{2,i} > N-\kappa\}}, \tag{3}$$

where \bar{C}_N denotes the empirical survival copula (Nelsen 2006, pp. 32–34), $\mathbb{1}_{\{\bullet\}}$ is the indicator function, $R_{1,i}$ and $R_{2,i}$ denote the ranks of $x_{1,i}$ and $x_{2,i}$, for $i = 1, \dots, N$, and κ a threshold rank to be chosen by using for instance the heuristic plateau-finding algorithm described by Schmidt (2003) (see also Frahm et al. 2005).

As far as λ_U^{CFG} is concerned, Frahm et al. (2005) showed that the underlying hypothesis of EV copula is not so stringent and that the estimator also performs quite well when the assumption is not fully matched. However they also highlighted that the estimator can be strongly biased when data are not tail dependent. Moreover, as for every asymptotic property, a reliable estimation of λ_U from a finite (and usually small) sample can be problematic (if not almost impossible). Since the claim of the good performance of λ_U^{CFG} even for non-EV copulas is sometimes used to justify its use, we performed a simulation study to better understand the actual magnitude of its bias and uncertainty for the whole range of positive correlation values (0, 1). We used four copulas (Gauss, Student with four degrees of freedom, Gumbel, and Frank). Gauss and Frank copulas exhibit zero tail dependence, whereas Gumbel and Student show positive UTD. For each model we have simulated 1,000 bivariate samples with size 1,000 and 20 different values of $\tau_K \in (0, 1)$. For each sample we have computed the value of λ_U^{CFG} and λ_U^{SS} . Since a visual assessment can be more effective than tables usually reported for this type of simulation studies, we have summarized the results by the $\lambda_U - \tau_K$ plane

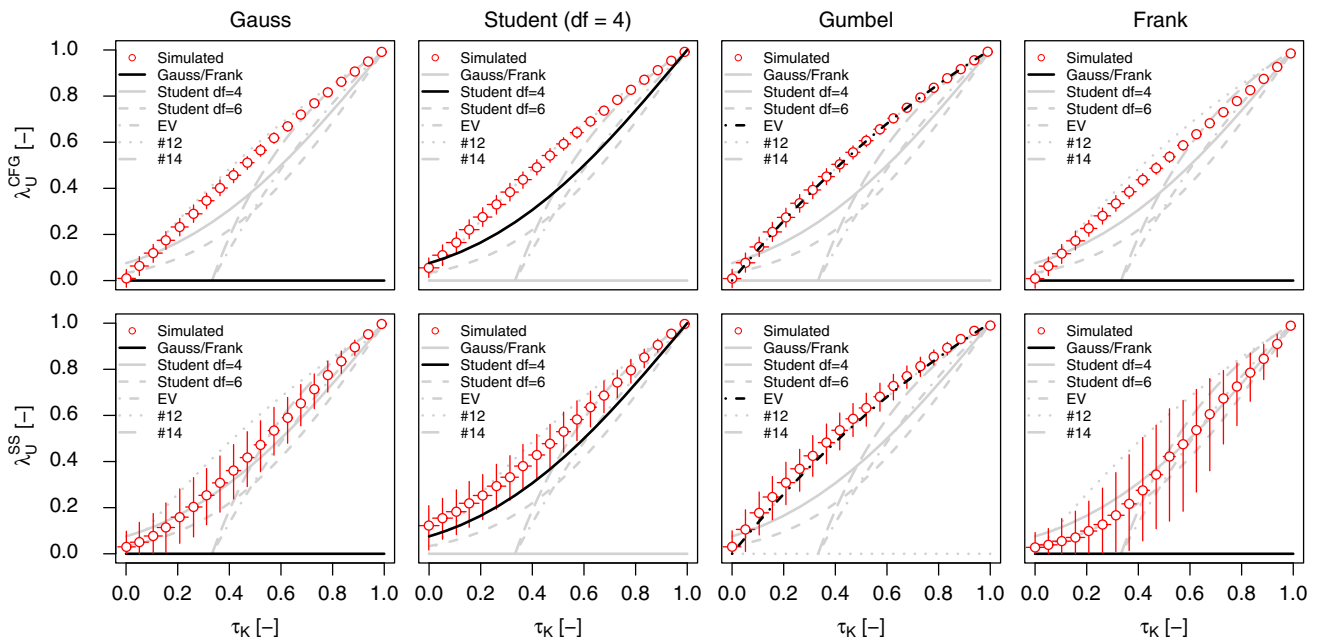


Fig. 1 $\lambda_U - \tau_K$ diagrams for four copula families (Gauss, Student with four degrees of freedom, Gumbel, and Frank). *Points* denote the average values over 1,000 simulated samples of size 1,000. *Crosses* denote the 95 % Monte Carlo confidence intervals of λ_U and τ_K

(Serinaldi 2008) shown in Fig. 1. The diagrams clearly show that λ_U^{CFG} always returns estimates close to the theoretical curve corresponding to Gumbel copula (the same curve holds for other EV copulas, such as Galambos, Hüsler–Reiss, and Tawn copulas) independently of the true underlying model. Moreover, considering the very small variance of λ_U^{CFG} , it is easy to conclude that there is a quasi one-to-one relationship between τ_K and λ_U estimates which is very close to the theoretical relationship that characterizes EV copulas, thus making the estimator almost uninformative: indeed, if the copula is EV we can use the theoretical relationship $\tau_K - \lambda_U$ (see e.g., Serinaldi 2008), whereas if the copula is not EV, λ_U^{CFG} gives estimates close to the theoretical curve in any case.

Unfortunately, λ_U^{SS} does not improve the situation very much. This estimator exhibits a slightly larger variability, but a persistent bias for Student, Gauss, and Frank copulas. In particular, λ_U^{SS} correctly returns zero values for Gauss and Frank copulas only for $\tau_K < 0.15$ and < 0.4 , respectively. Thus, also this estimator provides unreliable results for moderate and high values of the overall correlation, whose value dominates the potential values attainable for λ_U . It should be noted that the sample size 1,000 was chosen to be much larger than the sample size usually available in hydrology (i.e. 50–100) to show that relevant uncertainty also characterizes relatively large samples. Figure 2 shows that a small sample size further increases both bias and uncertainty, thus making the estimates

estimates. *Lines* indicate theoretical relationships corresponding to Gauss, Frank, Student with four degrees of freedom, Student with six degrees of freedom, EV copulas, and Archimedean copulas “#12” and “#14” as listed by Nelsen (2006, pp. 116–119)

uninformative. These results explain the behavior reported in Fig. 15 of Serinaldi (2008) and confirm the need for a thorough re-assessment of those analyses (which is done in the next sections) as well as all the studies whose conclusions are based on these estimators.

2.2 Formal statistical tests for upper tail independence

In order to overcome the shortcomings of the λ_U estimators, we need alternative tools. Following one of the recommendations of Frahm et al. (2005), we consider some formal statistical tests introduced by Falk and Michel (2006) and Zhang (2008). We briefly summarize the rationale of these methods referring to the original works for all details.

The tests proposed by Falk and Michel (2006) are based on the following findings. Let (X_1, X_2) a random vector with values in $(-\infty, 0]^2$, whose joint distribution function $G(x_1, x_2)$ coincides with a max-stable or EV distribution with reverse exponential marginals $G(x, 0) = G(0, x) = \exp(x)$, $x \leq 0$, where G has the form:

$$G(x_1, x_2) = \left(\exp(x_1 + x_2) D \left(\frac{x_1}{x_1 + x_2} \right) \right), \quad x_1, x_2 \leq 0, \tag{4}$$

in which $D : [0, 1] \rightarrow [1/2, 1]$ is the so-called Pickands dependence function (see Pickands III 1981; Falk and Reiss

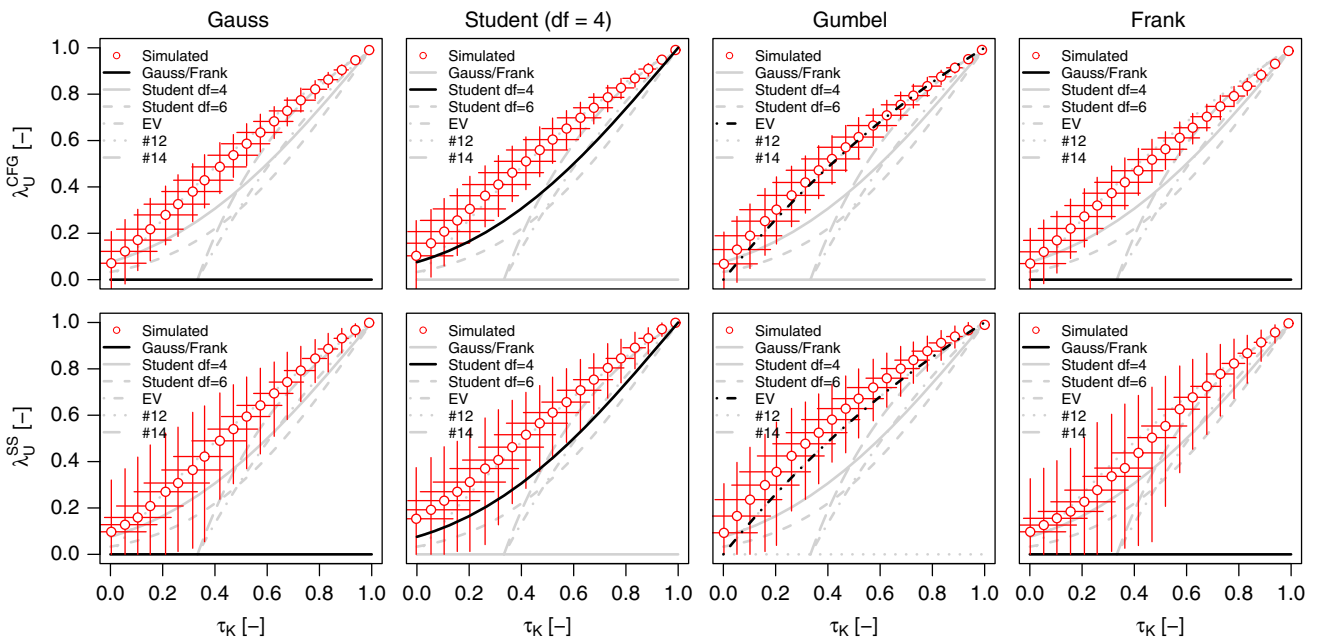


Fig. 2 As for Fig. 1 but for samples of size 75

2005 and references therein). Falk and Michel (2006) found that the conditional distribution of $X_1 + X_2$, given $X_1 + X_2 > c$ has an asymptotic distribution function $F(t) = t^2, t \in [0, 1]$, as $c \rightarrow 0^-$ if and only if X_1 and X_2 are tail independent, i.e. $D(z) = 1, z \in [0, 1]$. If D is not constant and equal to 1, then the asymptotic distribution is the standard uniform distribution $F(t) = t, t \in [0, 1]$. This property is used to define tests for tail independence derived from the Neyman–Pearson lemma as well as standard goodness-of-fit tests such as Fisher, Kolmogorov–Smirnov and χ^2 . It should be noted that the arbitrary marginals can be transformed into reverse exponential by simple (parametric or semi-parametric) quantile transformations $y = \log(F_X(x))$, where F_X can be the known distribution of X or a parametric distribution fitted on the observed values x_i , or the empirical distribution function of x_i (see e.g. Zhang 2008 for a description of the different options).

For a given sample $(x_{1,1}, x_{2,1}), \dots, (x_{1,N}, x_{2,N})$ of (X_1, X_2) (with reverse exponential marginals), choose $c < 0$ and select the observations $x_{1,i} + x_{2,i}$ such that $x_{1,i} + x_{2,i} > c$ and denote them as $C_1, \dots, C_{K(N)}$. Falk and Michel (2006) showed that $\mathcal{V}_i = C_i/c$ are independent and identically distributed with distribution F_c , if c is large enough, and they are independent of $K(N)$, which in turn is binomial distributed $B(N, q)$ with $q = 1 - (1 - c) \exp(c)$. Therefore we have to test the null hypothesis $F_c = t^2$ (which holds for tail independence) against the alternative $F_c = t$ (which holds for tail dependence). Under the above assumptions the optimal test is based on the log-likelihood ratio:

$$T(\mathcal{V}_1, \dots, \mathcal{V}_m) := \log \left(\prod_{i=1}^m \frac{1}{2\mathcal{V}_i} \right) \quad m = K(N) > 0. \quad (5)$$

If m is large enough, for the central limit theorem, the p -values of the optimal test derived from the Neyman–Pearson lemma is

$$p_{NP} \approx \Phi \left(\frac{2 \sum_{i=1}^m \log(\mathcal{V}_i) + m}{m^{1/2}} \right), \quad (6)$$

where Φ denotes the standard Gauss cumulative distribution function.

The other tests proposed by Falk and Michel (2006) rely again on C_i/c through the variables

$$\mathcal{U}_i := F_c(C_i/c) = \frac{1 - (1 - C_i) \exp(C_i)}{1 - (1 - c) \exp(c)}, \quad i = 1, \dots, m. \quad (7)$$

Under the null hypothesis, \mathcal{U}_i are independent and uniformly distributed on $(0, 1)$. Thus, we simply need to test this hypothesis applying some goodness-of-fit tests in a standard way. Falk and Michel (2006) suggested Fisher, Kolmogorov–Smirnov and χ^2 tests. These tests are available in R (R Core Team 2013) by the POT package (Ribatet 2006). Falk and Michel (2006) showed that the Neyman–Pearson test has the smallest type II error rate, followed by Kolmogorov–Smirnov and χ^2 , whereas Fisher exhibits a poor performance. On the other hand, Neyman–Pearson does not control the type I error if the value of the threshold c is too far from 0, whereas the other tests control the type I error rate for any c .

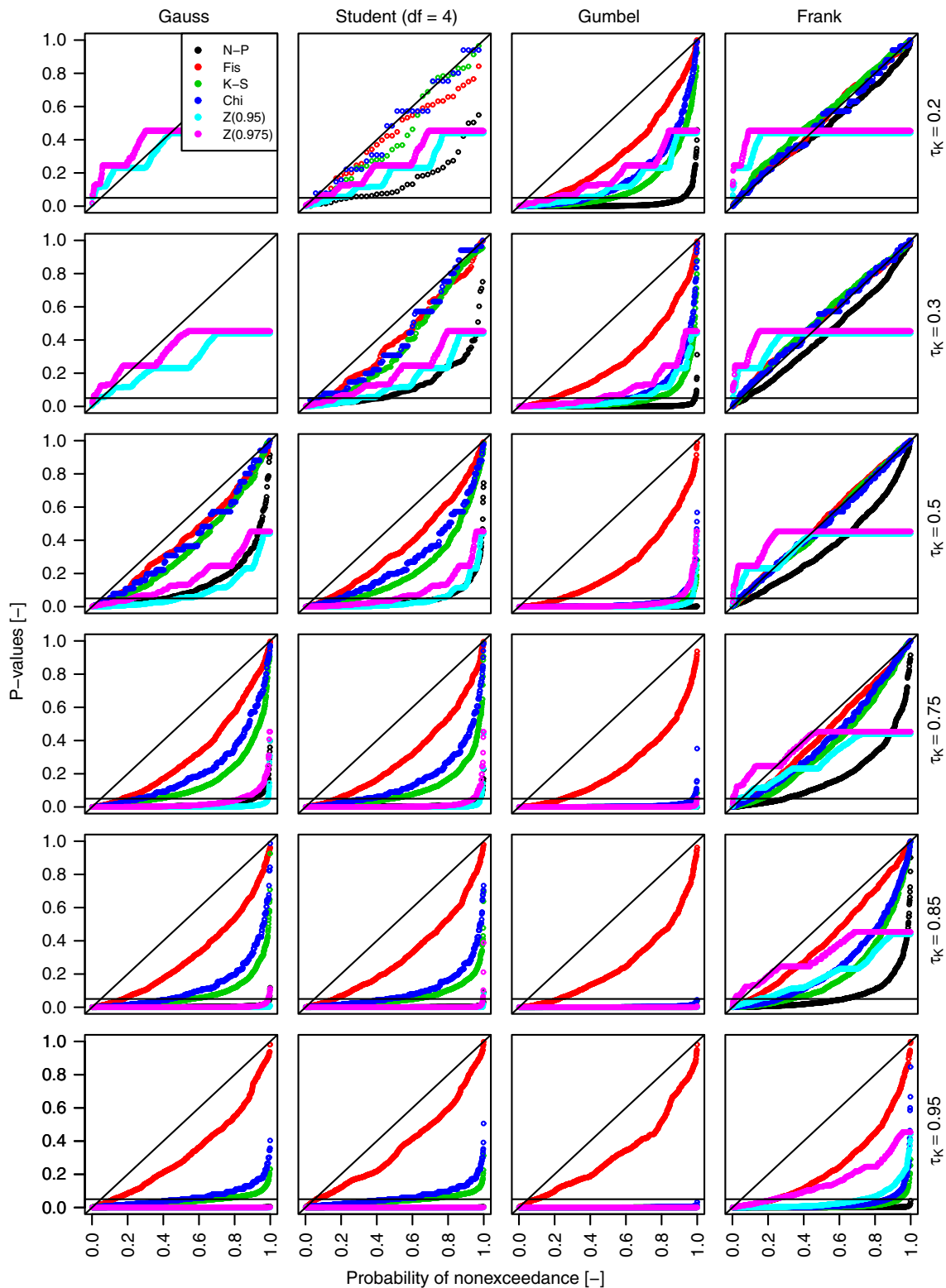


Fig. 3 Distributions of the p -values resulting from 1,000 testing exercises performed on bivariate samples of size 1,000 drawn from four models accounting for the full range of possible positive dependence (as measured by τ_K). Five formal tests for upper tail independence are considered: Neyman–Pearson (“N–P”), Fisher (“Fis”), Kolmogorov–Smirnov (“K–S”), χ^2 (“Chi”), and Zhang

with $p = 0.95$ and $p = 0.975$ and (“Z(0.95)” and “Z(0.975)”). Note that for Gauss copula and the smallest values of τ_K , four tests (N–P, Fis, K–S, and Chi) provide no outcomes as the data over the threshold c are not enough to compute test statistics. See text for further details about settings and interpretation

The test for upper tail independence proposed by Zhang (2008) is based on the so-called tail quotient correlation defined as

$$\rho_Q := \frac{\max_{i \leq N}(\mu + w_{2,i})/(\mu + w_{1,i}) + \max_{i \leq N}(\mu + w_{1,i})/(\mu + w_{2,i}) - 2}{\max_{i \leq N}(\mu + w_{2,i})/(\mu + w_{1,i}) \times \max_{i \leq N}(\mu + w_{1,i})/(\mu + w_{2,i}) - 1}, \tag{8}$$

mance already recognized by Falk and Michel (2006). Since this test cannot discriminate between tail dependent and tail independent models irrespective of the correlation

where μ is a positive threshold, $w_{1,i}$ and $w_{2,i}$ are exceedance values over the threshold μ of positive random variables X_1 and X_2 . Zhang (2008) assumed that X_1 and X_2 are identically distributed with unit Fréchet distribution function $F_X(x) = \exp(1/x)$, $x > 0$. When the variables do not fulfill this condition, they can be easily transformed as for the other tests (see Zhang 2008 for a detailed discussion). In this study, we use the transformation $y = -\log(\hat{F}_X(x))$, where $\hat{F}_X = \frac{1}{N+1} \sum_{i=1}^N \mathbb{1}_{\{X < x\}}$. Zhang (2008) showed that under the hypothesis that X_1 and X_2 are tail independent, for $N \rightarrow \infty$, ρ_Q is asymptotically gamma distributed $\Gamma(N\rho_Q; 2, 1 - e^{-1/\mu})$.

In order to check the finite-sample behavior of the tests described above, we have performed a Monte Carlo experiment applying the same settings used to study the properties of the λ_U estimators. Building on the results of Falk and Michel (2006), the threshold c was set equal to -0.1 , whereas the Zhang test was performed using two p quantile threshold values $\mu = x_p$, $p = \{0.95, 0.975\}$. Similarly to λ_U analysis, we explored the impact of the overall correlation (measured by τ_K) on the UTD recognition. Results are shown in Fig. 3 by probability plots. The diagram must be interpreted as follows: under the assumption that the null hypothesis is true (i.e., tail independence) the p -values resulting from the 1,000 experiments are expected to be uniformly distributed on $(0,1)$ and then be aligned along the 1:1 line; if the alternative is true (i.e., tail dependence), and a test is performed at the 5 % significance level, we expect to observe that 95 % of p -values are smaller than 0.05; results in the middle provide information about the actual power of the tests.

Focusing on the Gauss copula (with known theoretical tail independence), we expect p -values aligned along the 1:1 line. Similar to λ_U , results depend on the strength of the overall correlation. As τ_K increases the rejection rate increases. For the Gauss copula, Neyman–Pearson and Zhang tests give the highest (incorrect) rejection rate for high τ_K values, whereas Kolmogorov–Smirnov and χ^2 tests exhibit a rejection rate around 0.5 even for high correlation values. Results for Fisher test confirm the poor perfor-

value, it is not further discussed in the following. To make a fairer comparison, it is worth focusing on the patterns corresponding with a specific correlation value. For $\tau_K = 0.5$, Kolmogorov–Smirnov and χ^2 tests show the patterns closest to 1:1 line (and a rejection rate close to 5 %), whereas Neyman–Pearson tends to over-rejection. The performance of Zhang test depends on the choice of the threshold μ (as expected), the performance improving as the threshold increases. Note that the p -values of Zhang test are computed up to 0.4 as we observed that the asymptotic Γ distribution gives a good approximation only in the upper tail. However, this is not a true shortcoming as we are interested in the smallest p -values.

For the Frank copula, we expect in principle results similar to Gauss. However, the shape of the upper tail of the Frank copula is rather different from that of Gauss, showing a wider spread. This explains why almost all tests give good results for $\tau_K = 0.5$.

As far as Student is concerned, we used four degrees of freedom because this value returns copulas with strong UTD close to EV copulas (see theoretical curves in Fig. 1). The second column of plots in Fig. 3 show that Neyman–Pearson and Zhang with $\mu = x_{0.95}$ return the best results (among the competitors) with a rejection rate close to 60 % for $\tau_K = 0.5$. As τ_K increases, their performance improves reaching values close to the expected 95 % (assuming that we perform the test at the 5 % significance level). For Student copula, Kolmogorov–Smirnov and χ^2 tests show quite a low rejection rate.

Moving to the true EV Gumbel copula, all tests perform very well for $\tau_K = 0.5$. Of course, the performance deteriorates as τ_K decreases and the dependence structure approaches the product copula (overall independence). However, the difference between the EV Gumbel copula and Student is remarkable, thus indicating that the considered tests are sensitive not only to UTD by itself but also to the nature of the overall structure of dependence and the shape of the joint upper tail. In this respect, it is worth noting that the tests proposed by Falk and Michel (2006) were tested only on EV models with $\tau_K \in \{0.33, 0.5\}$ or

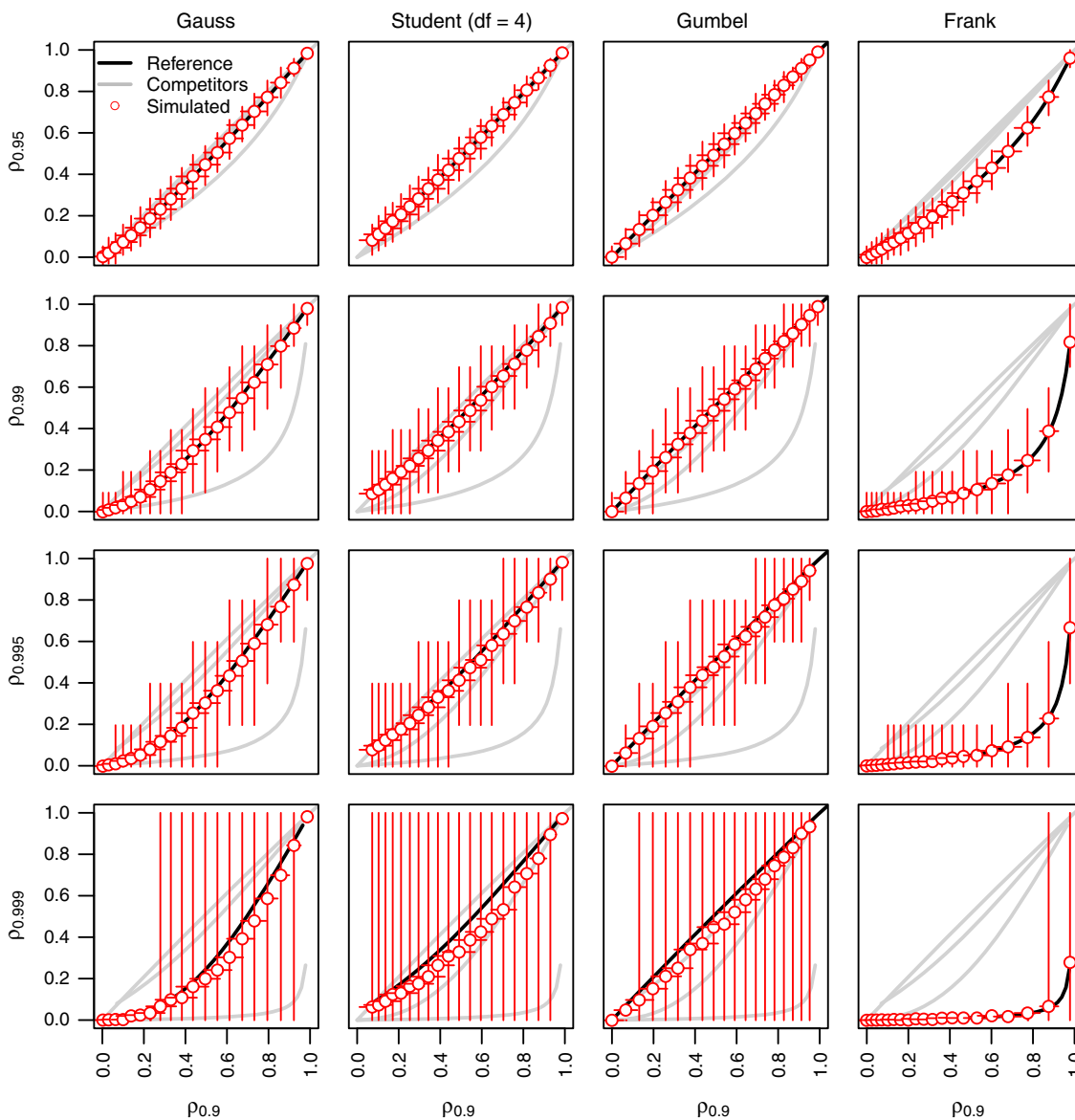


Fig. 4 $\rho_{11} - \rho_{12}$ diagrams for four copula families. Points denote the average values over 1,000 simulated samples of size 1,000. Crosses denote the 95 % Monte Carlo confidence intervals of ρ_t estimates. Each point (between 0 and 1) of the theoretical curves and each point

correspond with different values of $\tau_K \in (0, 1)$ according to a monotonic increasing relationship (i.e. $\rho_t = 0$ for $\tau_K = 0$ and $\rho_t = 1$ for $\tau_K = 1$)

complete independence. Our results show that these are ideal conditions for these tests, whereas for non-EV and tail independent models we obtain intermediate results when the overall correlation is not null and void. Since the true structure of dependence is commonly unknown in real-world problems, the above results are fundamental to avoid misleading conclusions, confusing for instance strong dependence of upper quantiles resulting from mid-high overall correlation with true UTD.

Finally, we performed simulation experiments also for smaller sample sizes (results not shown). In these cases the

performance of all tests rapidly worsens and often no results are returned because the data are not enough to apply a suitably high threshold. It is worth further stressing that both λ_{UJ} and the formal tests above attempt to measure/recognize asymptotic properties and rely on limiting distributions/properties which hold for $N \rightarrow \infty$. Therefore, trying to use such tools on small samples such as less than 100 annual maxima without additional information (e.g., Serinaldi et al. 2009; Janga Reddy and Singh 2014) is essentially a speculation exercise such as inferring the EV nature of the underlying dependence structure.

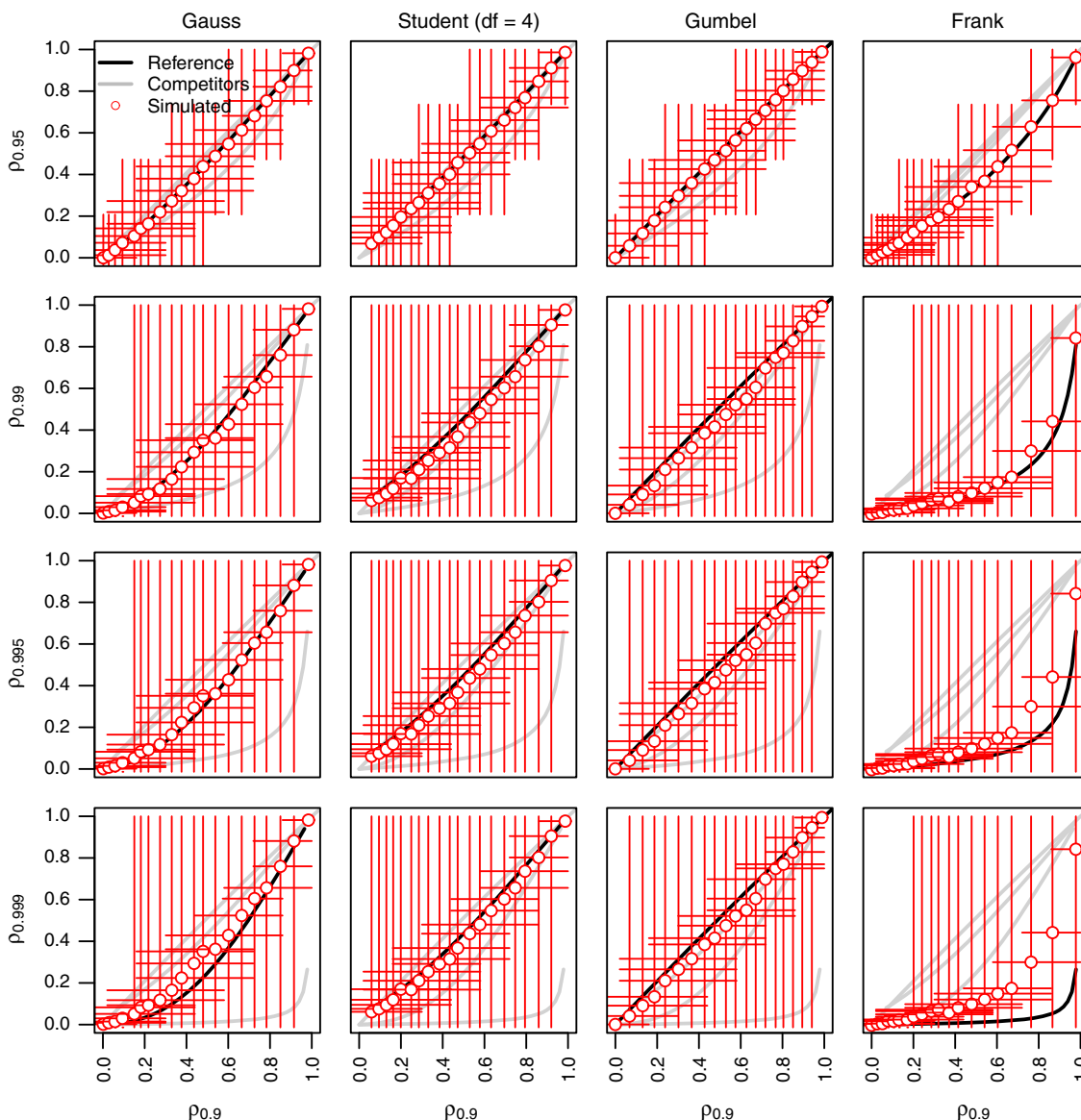


Fig. 5 As for Fig. 4 but for samples of size 75

2.3 Pairwise binary correlation and binary entropy on triples

After studying the finite sample properties of diagnostics devised for asymptotic pairwise UTD, we introduce some diagnostics that focus on the upper tail but do not resort to limiting properties. The aim is to use indices based on the available data rather than asymptotic results, easy to compute and apply, and allowing for effective visualization of high-dimensional data sets. In this respect, as a first option, we consider the pairwise correlation of binary vectors ρ , where the binary vectors describe the occurrence of u quantile threshold exceedances. Thus, ρ is defined as

$$\begin{aligned} \rho(x_1, x_2) &= \frac{\mathbb{P}[X_1 > x_1 \cap X_2 > x_2] - \mathbb{P}[X_1 > x_1]\mathbb{P}[X_2 > x_2]}{\sqrt{\mathbb{P}[X_1 > x_1](1 - \mathbb{P}[X_1 > x_1])\mathbb{P}[X_2 > x_2](1 - \mathbb{P}[X_2 > x_2])}} \\ &= \frac{C(u_1, u_2) - u_1 u_2}{\sqrt{u_1(1 - u_1)u_2(1 - u_2)}}. \end{aligned} \tag{9}$$

For $u_1 = u_2 = t$, we have

$$\rho_t := \rho(t, t) = \frac{C(t, t) - t^2}{t(1 - t)}. \tag{10}$$

By choosing a suitable set of t values, we can build diagrams of ρ_{t_1} versus ρ_{t_2} . In this study, we use $t_1 \in$

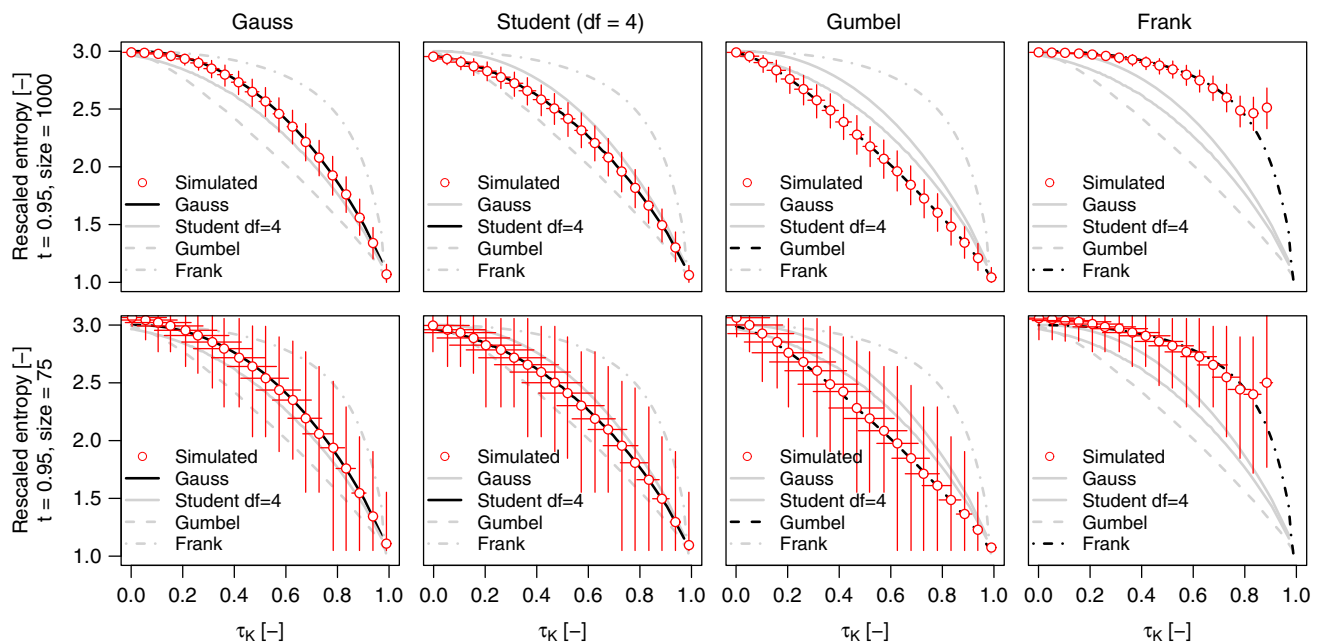


Fig. 6 $S - \tau_K$ diagrams for four copula families (Gauss, Student with four degrees of freedom, Gumbel, and Frank). Points denote the average values over 1,000 simulated samples of size 1,000 (top) and 75 (bottom). Crosses denote the 95 % Monte Carlo confidence

intervals of S and τ_K estimates. Lines indicate theoretical relationships corresponding to Gauss, Student with four degrees of freedom, Gumbel, and Frank copulas. S is rescaled by the factor $-1/(t \log_2(t) + (1-t) \log_2(1-t))$

$\{0.95, 0.99, 0.995, 0.999\}$ and $t_2 = 0.9$ for the sake of illustration, being however possible to draw $\rho_{t_1} - \rho_{t_2}$ diagrams for whatever pair of values (t_1, t_2) . The reliability of these diagrams is tested via the same simulation experiments used for checking the performance of λ_U and formal tests for upper tail independence. Results are reported in Fig. 4, where each point (between 0 and 1) of the theoretical curves and each point indicating the average values of ρ_{t_1} and ρ_{t_2} over 1,000 simulated samples correspond with different values of $\tau_K \in (0, 1)$ according to a monotonic increasing relationship (i.e. $\rho_t = 0$ for $\tau_K = 0$ and $\rho_t = 1$ for $\tau_K = 1$). Compared with λ_U estimators, ρ_t is almost unbiased, a small bias emerging for very high quantile thresholds (especially for Student and Gumbel models). However, in these cases, the 95 % confidence intervals cover the whole range of possible values (0, 1), thus making very difficult to discriminate between different competitors. The only exception is the Frank copula, whose upper tail has a shape markedly different from the other models, which on the contrary behave similarly, especially when the analysis relies on finite samples. For the sake of completeness, Fig. 5 shows the results for samples of size 75. In this case, we have some bias for high thresholds and Frank copula. However, the most important aspect is the high uncertainty of the estimates that makes any conclusion on the upper tail extremely difficult, further stressing

the low (or almost null and void) reliability of the inference concerning the behavior of the tails when the analysis is based on small samples (without additional information).

Finally, we consider the binary entropy on triples introduced by Bárdossy and Pegram (2009). The rationale of this index is to overcome a pairwise assessment in order to look effectively at the high-order dependence properties. This measure can be applied to any triple of variables, but for the sake of simplicity and without loss of generality, let us consider three variables (e.g., rainfall or stream flow records) recorded at three different locations. Fix a quantile threshold for each time series and define the corresponding binary vectors as for ρ_t , assigning for instance the value 1 to the records exceeding the threshold and 0 otherwise. At each time, the state of a triple is described by the set $\{i, j, k\}$, for $i, j, k = 0, 1$. Thus, if all three locations are under threshold we have $\{0, 0, 0\}$, whereas if they are over threshold we have the state $\{1, 1, 1\}$. For each triple we have $2^3 = 8$ possible mutually exclusive states (i.e. the number of possible permutations of length three from an alphabet of two symbols $\{0, 1\}$), and eight binary probabilities $p(i, j, k)$, for $i, j, k = 0, 1$, can be calculated over the N realizations, recalling that the states 0 and 1 are the lower and upper partition of the probabilities by the t quantile threshold. Therefore, for example, the probability that all three locations are simultaneously under threshold

or over threshold are $p(0, 0, 0)$ and $p(1, 1, 1)$, respectively. The information entropy S (Shannon 1948) of each of the sets of eight probabilities is calculated as a measure of dependence in a given triple; thus

$$S = - \sum_{i,j,k=0}^1 p(i, j, k) \log_2(p(i, j, k)). \tag{11}$$

For ease of interpretation, we use the normalized Shannon entropy $\mathcal{H}_S := S/S_{\max} \in [0, 1]$, where $S_{\max} = \log_2(8)$ is the maximum value of S corresponding with the set of probabilities $P_e = \{1/8, \dots, 1/8\}$, describing the uniform distribution. Independence corresponds with the set of probabilities $\{(1-t)^3, (1-t)^2t, (1-t)t^2, (1-t)^2t, (1-t)t^2, (1-t)t^2, (1-t)t^2, t^3\}$, whereas comonotonicity with the condition that $p(0, 0, 0) = (1-t)$, $p(1, 1, 1) = t$, and the other probabilities equal to zero.

Since \mathcal{H}_S implies triples whose mutual correlations can be rather different, general diagnostic plots such as $\lambda_U - \tau_K$ or $\rho_{i1} - \rho_{i2}$ diagrams cannot be defined. Nevertheless, setting $\tau_K(x_1, x_2) = \tau_K(x_1, x_3) = \tau_K(x_2, x_3)$ and $t = 0.95$, Fig. 6 shows the performance of the S estimator for samples of size 1,000 and 75 drawn from four different dependence structures. Similar to ρ_i , entropy on triples is almost unbiased and uncertainty increases as the sample size decreases and $t \rightarrow 1$.

For empirical analyses, the comparison between observed and theoretical dependence structures is carried out by ad hoc simulations. In particular, in order to fully preserve the marginal distributions of the records and highlight only the effect of the dependence structure, we

have applied the following simulation-supported rank-permutation approach:

1. For each triple of time series $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ with size N , the observed records in each vector \mathbf{x}_i , $i = 1, 2, 3$, are replaced by their ranks $(\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3)$; identical values (statistical ties) are handled by randomization. Note that in this context, randomization does not affect the final results.
2. For each triple, simulate a sample with size N from a three-variate distribution function and replace the simulated values $(\mathbf{x}_1^s, \mathbf{x}_2^s, \mathbf{x}_3^s)$ with their ranks $(\mathbf{R}_1^s, \mathbf{R}_2^s, \mathbf{R}_3^s)$. In this study we use Gauss and Student (with four degrees of freedom) as they allow us to assess the impact of tail dependence by models that preserve pairwise mutual correlations (embedded in the correlation matrix).
3. Replace the simulated (cross-correlated) vectors of ranks with the observations whose observed ranks match the simulated ranks, i.e. $(\mathbf{x}_1(\mathbf{R}_1^s), \mathbf{x}_2(\mathbf{R}_2^s), \mathbf{x}_3(\mathbf{R}_3^s))$.

This procedure generates samples with marginal distributions identical to the observed and a desired latent dependence structure (here, Gauss and Student).

When the variables studied are spatially distributed and the dependence structure reflects the nature of the spatial organization, the mutual position of the triples is important (Bárdossy and Pegram 2009). Indeed, the triangle cannot be too long and thin, else there will be two points close together or one near the middle of a line joining the other two. Even though the ideal is an equilateral triangle, in

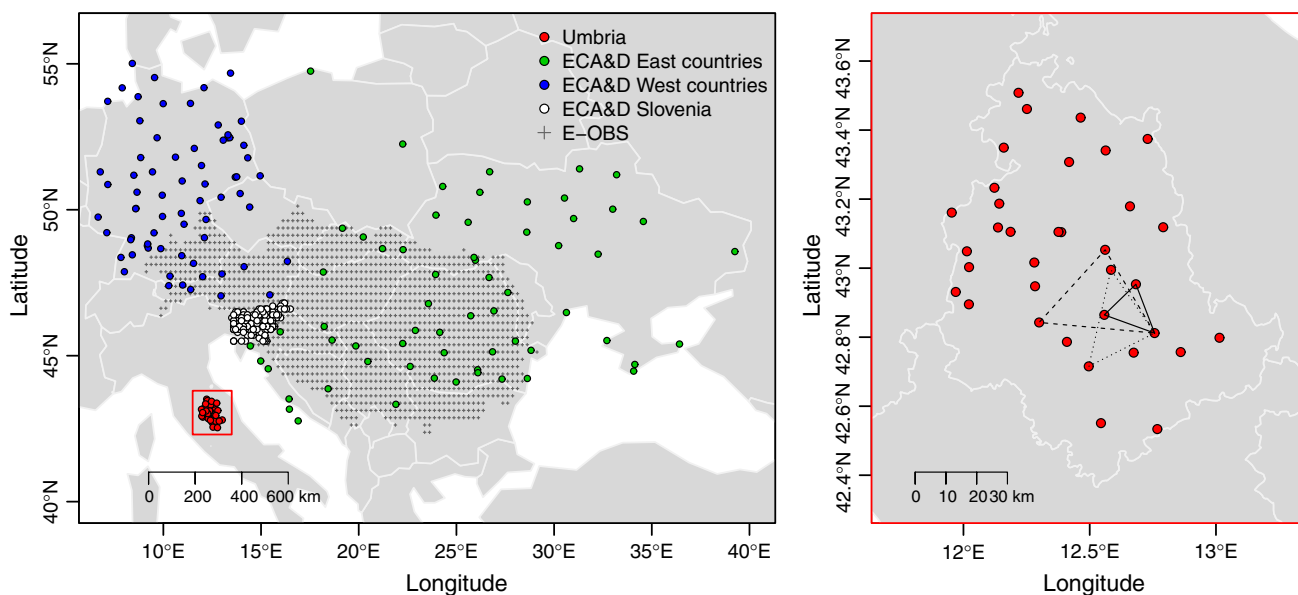


Fig. 7 Maps of rainfall stations and grid points of the data sets used in the empirical analyses. ECA&D data are distinguished between East countries, West countries and Slovenia as is specified in Sect.

4.1. Three examples of triples of sites suitable to compute \mathcal{H}_S are shown in the *right*

randomly scattered sites a compromise is necessary. As we used Heron's formula for the calculation of the triangle area A , the acceptance of a suitable triple of points relies on the following criterion (Bárdossy and Pegram 2009): for each of the three pairs of sides in an adopted triangle, we chose that the maximum difference in a pair must be less than 10 % of the perimeter of the triangle; i.e. for sides s_1 , s_2 , and s_3 , $\mathcal{P} = s_1 + s_2 + s_3$, the criterion is: accept triple if $\max |s_i - s_j|/\mathcal{P} < 0.1$, $\forall i \neq j$. Once the triangle has been identified, A is calculated. Diagrams of \mathcal{H}_S versus A are used as diagnostics of the relationship between the areal information content and the spatial scale.

3 Data

As mentioned in Sect. 2.1, we first re-evaluate the properties of the rainfall data studied by Serinaldi (2008). Referring to that work for further details, we recall that the data set comprises 35 rainfall series collected from 1995 to 2001 at 30-min time scale by a network of raingauges

located in Umbria (central Italy). Ten time series are complete whereas the others show some intervals of missing data in 1995 and 1996. However, this does not affect the analyses very much, as they focus on the spatial correlations between simultaneous observations. In order to study the behavior of the rainfall fields at different time scales, the data were aggregated at 1, 3, 6, 12, and 24 h. The mutual distances between stations range from 0.9 to 120 km (see Fig. 7).

Since good-quality data sets at a fine temporal scale are available only for relatively small areas and a few years, in order to study the spatial structure of rainfall and draw general results, the data set above was complemented with 287 daily rainfall series extracted from the ECA&D database available at the web site <http://eca.knmi.nl/dailydata/predefinedseries.php> (Klein Tank et al. 2002). These time series cover more than 40 years with less than 10 % of missing values. They are shifted in time, but the length ensures that simultaneous observations to be used in correlation analyses are close to 40 years. This data set allows us to expand the analysis on super-daily time scales (namely, 2,

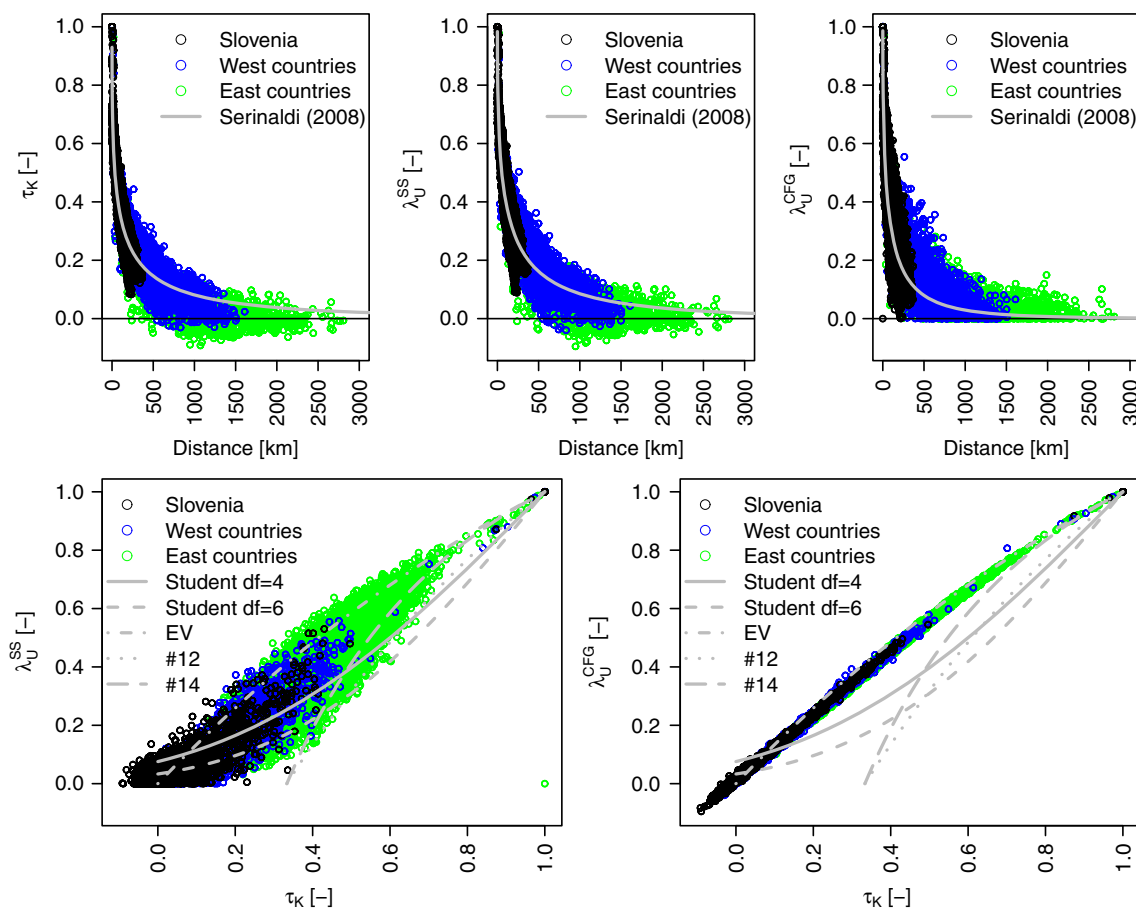


Fig. 8 Relationship between λ_U and τ_K and inter-station distance, and $\lambda_U - \tau_K$ diagrams for central-eastern Europe daily data

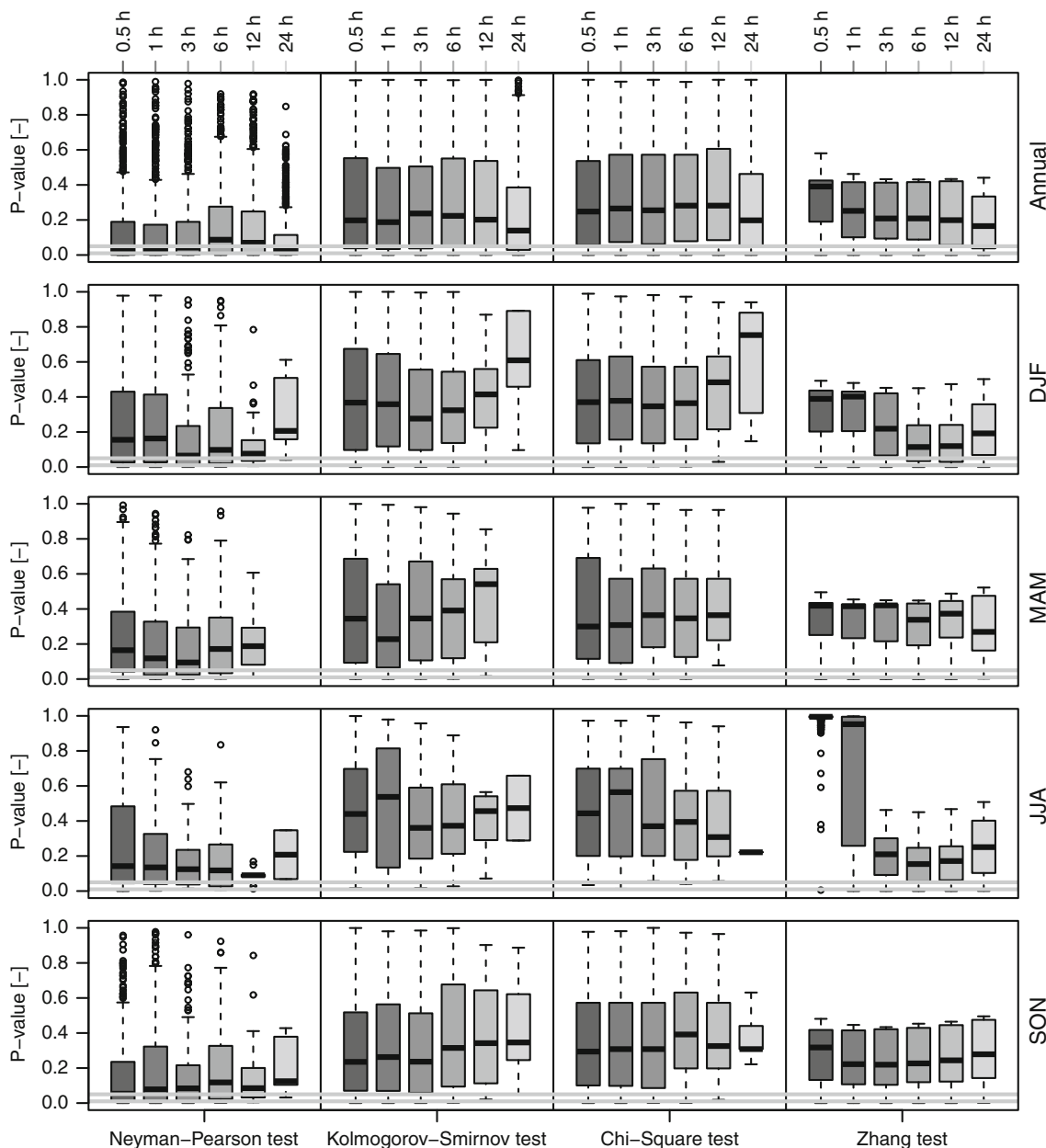


Fig. 9 UTD test results for Umbria rainfall data

5, 10, and 30 days) and wider spatial scales (up to a mutual distance between stations of $\approx 3,000$ km; see Fig. 7).

Some analyses were performed on an additional data set already studied by Serinaldi and Kilsby (2014a), that comprises $0.25^\circ \times 0.25^\circ$ (i.e., $\approx 25 \times 25$ km²) rainfall gridded data covering the Danube basin (1,462 grid points; Fig. 7) extracted from the E-OBS database developed within the EU-FP6 project ENSEMBLES and available at the web site <http://eca.knmi.nl/download/ensembles/download.php> (Haylock et al. 2008). The selected data set covers the 32-year period between 1950 and 1981, in which the

data show a small number of missing values and reasonably homogeneous coverage over the entire area.

4 Empirical results

4.1 Results for λ_U estimators

For the Umbria data set, Serinaldi (2008) provided a detailed analysis of the relationship between λ_U and inter-station distance, and $\tau_K - \lambda_U$ diagrams. Similar patterns of

λ_U versus distance were also found by AghaKouchak et al. (2013). However, since the results reported in Sect. 2.1 show that the λ_U estimators are generally biased and closely reflect the strength of the overall correlation of data, the above mentioned patterns are expected and do not add much information compared to the spatial patterns of τ_K . On the contrary, they indicate UTD even if it could not exist.

For the sake of completeness we report the same diagrams for the 287 daily time series of central-eastern Europe data set. Figure 8 confirms that the relationships λ_U -distance are almost insensitive to the spatial and temporal time scales. Indeed, the exponentiated power functions estimated by Serinaldi (2008) on the 30-min Umbria rainfall data fit the central-eastern Europe data well. The $\lambda_U - \tau_K$ diagrams also confirm the bias towards EV behavior. Schmidt and Stadtmüller (2006) already recognized that “an increasing correlation [ρ_p] deteriorates the results because of the increasing bias of the non-parametric λ_U estimator”. Note that the Europe data set was split in three subsets (western and eastern countries, and Slovenia; see Fig. 7) to avoid the concealing effect caused by the cluster of Slovenia’s stations on the pairwise calculations, and to show the coherence across the geographic area. Similar results can be found for the gridded data. The $\lambda_U - \tau_K$ diagrams look like those obtained for the simulated samples (Fig. 1) and further stress the strong bias of both λ_U^{CFG} and λ_U^{SS} estimators. The application of additional diagnostics is therefore mandatory.

4.2 Results for formal statistical tests for upper tail independence

Based on the power study described in Sect. 2.2, we discarded the Fisher test, and performed the Zhang test with $\mu = x_{0.975}$. In order to obtain a comprehensive picture, the tests were applied to the Umbria data set both in an annual and seasonal basis (distinguishing four seasons: DJF, MAM, JJA, SON) as the summer events are often convective and different tail dependence behavior can be expected. Moreover, we considered six time scales (from 30 min to 24 h) to explore the possible existence of a characteristic scale for which the rainfall dynamics generate UTD. With the same rationale, the central-eastern Europe data set was analysed both in an annual basis and extracting the summer months (approximately identified with MJJA). Daily data were aggregated up to 30 days. Finally the gridded data set (in an annual basis) was also considered in this analysis to understand the effect of spatial averaging (from $0.25^\circ \times 0.25^\circ$ up to $2^\circ \times 2^\circ$). Results are summarized as box-plots of the p -values resulting from each test performed pairwise. If UTD is a dominant property in rainfall fields, we expect a large number of small p -values (for instance, smaller than 0.05, if

we perform the test at the 5 % significance level). Of course, since this is a multiple testing exercise, a percentage of false positive results is expected. However, this aspect does not matter very much, as we anticipate that all results point to no much evidence for UTD, apart from cases corresponding to strong overall (τ_K) dependence and small inter-site distances.

Results for Umbria rainfall data in Fig. 9 show that only the Neyman–Pearson test returns a reasonable percentage of p -values ($\approx 50\%$) in between 0.01 and 0.05 for annual and SON data; however, it should be noted that in several cases the tests provide no outcomes as the data over the threshold c are not enough. This explains the lack of results for MAM season. In these cases every conclusion is speculative, but the lack of concurrent observations in the upper tail is reasonably an indication of lack of UTD. Therefore the overall conclusion is that if UTD does exist, it is not supported by empirical evidence (based on the set of formal tests used in this study). Analogous results hold true for the central-eastern Europe daily raingauge data and gridded data at all space and time scales mentioned above.

4.3 Pairwise binary correlation and binary entropy on triples

To further investigate the behavior of the UTD without resorting to asymptotic concepts, Fig. 10 shows the $\rho_{t1} - \rho_{t2}$ diagrams of binary correlation for several quantiles and temporal scales of aggregation for the Umbria data set. The patterns of JJA rainfall can be clearly distinguished from those of the other seasons for time scales below 12 hours, and indicate UTD stronger than that of rest of the year. This can easily be ascribed to the convective nature of the summer rainfall in this region (note that some time scales are not shown for the sake of better visualization). However, the main result is that the ρ_t ratios tend to lie below the Gaussian theoretical curve, thus denoting an upper tail correlation weaker than Gaussian. The cloud of points is close to the EV theoretical curve only for the JJA data, the smallest time scales (0.5 and 1 hour) and moderately extreme quantiles (i.e., $x_{0.95}$ and $x_{0.99}$). However, in these cases distinguishing between Gauss and EV behavior is difficult because the theoretical curves of the two models are too close to each other and confidence bands (not reported for the sake of readability) cover both curves. For the highest quantiles, the difference between the theoretical curves becomes more remarkable; however, the uncertainty increases (as the size of samples used to compute $\rho_{0.995}$ and $\rho_{0.999}$ decreases). Moreover, for the highest quantiles, a sub-Gaussian behavior is more evident.

Since the length of Umbria data set (seven years) is not enough to explore daily and coarser time scales, the diagrams in Fig. 10 are virtually extended using the gridded data. Figure 11 shows the results corresponding to

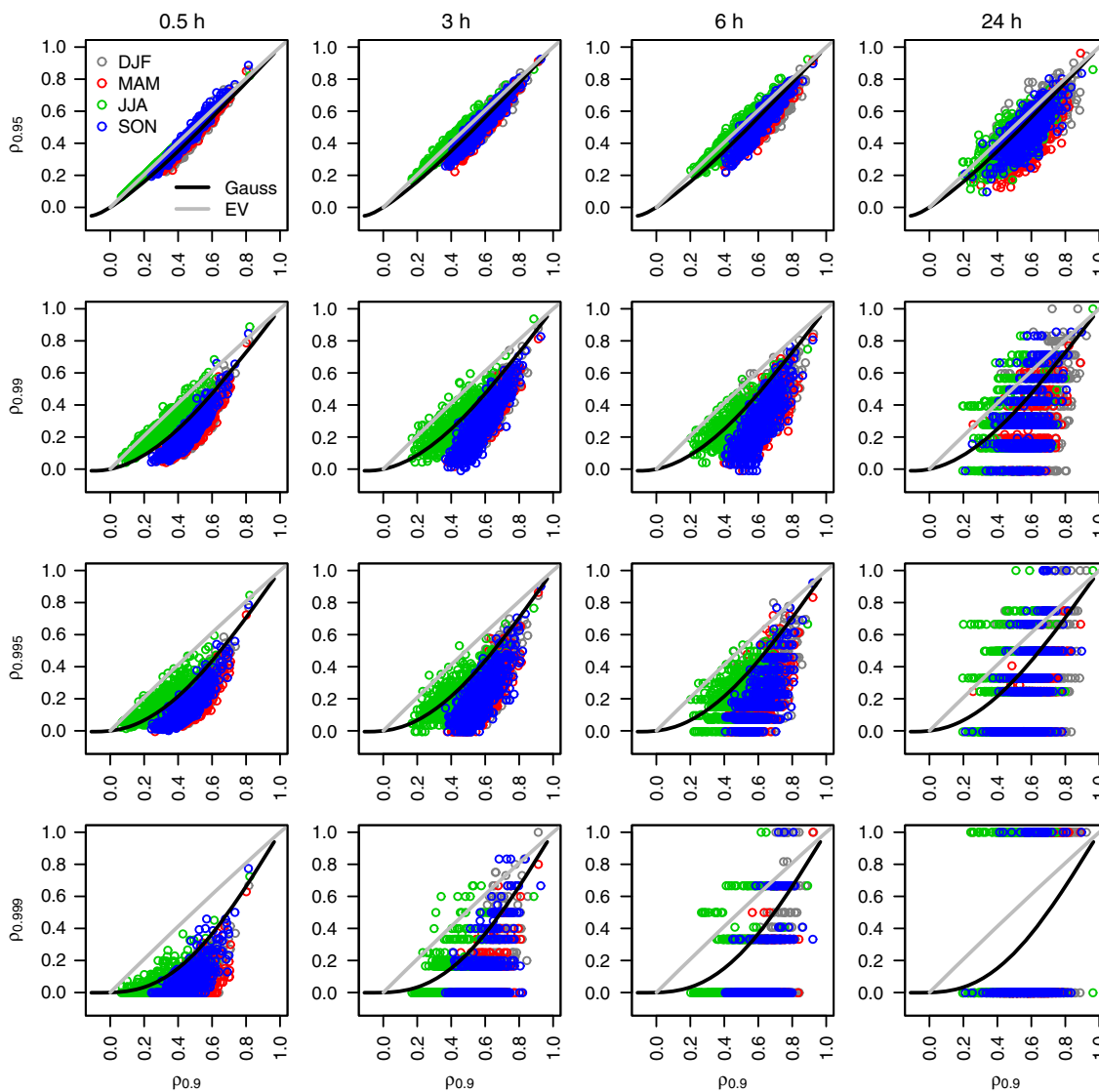


Fig. 10 $\rho_{11} - \rho_{12}$ diagrams for Umbria rainfall data

$0.25^\circ \times 0.25^\circ$ spatial resolution (similar results were found for the other spatial scales up to $2^\circ \times 2^\circ$). The patterns are very similar to those obtained for sub-daily scales. For the finest resolution (here, 1 day) and the largest sample size, the cloud of points is close to both EV and Gauss when the curves are close to each other, and to Gauss when the differences between the theoretical curves increase (i.e. for the highest quantiles).

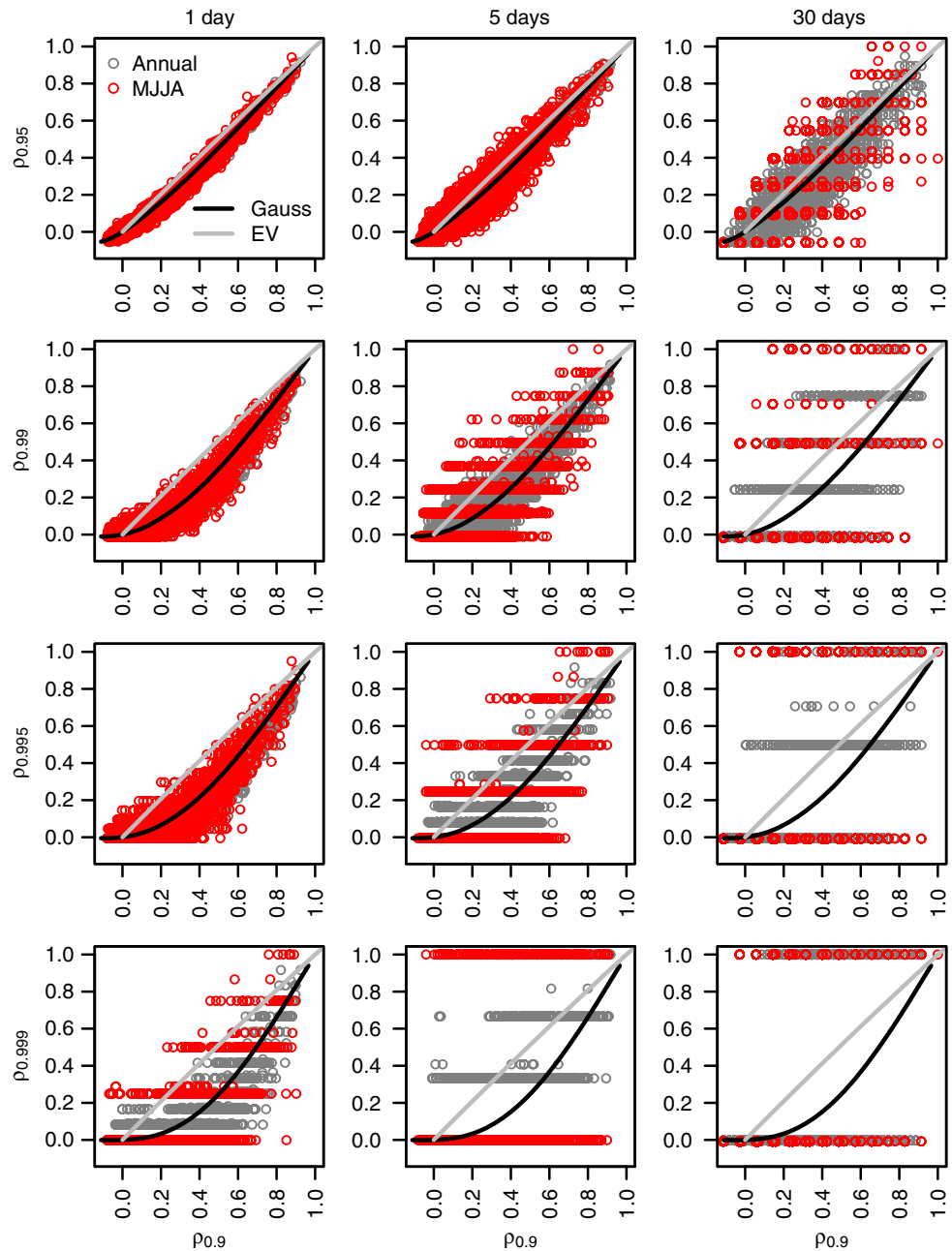
Therefore, the $\rho_{11} - \rho_{12}$ diagrams seem to confirm the results of the formal tests: when ρ_i ratios indicate possible UTD, the uncertainty and/or the small difference between EV and Gauss patterns do not allow us to distinguish, and when the difference is more evident, results indicate UTD close to or weaker than Gaussian.

However, is the Gaussian dependence structure enough to explain the spatial structure of the rainfall extremes? Our

working hypothesis is that hidden properties could be related to higher-order correlations. In this respect, \mathcal{H}_S can be a useful diagnostic. Following the procedures introduced in Sect. 2.3, we calculated \mathcal{H}_S on 443 triples describing approximately equilateral triangles for the Umbria data set, and for the resampled data with latent Gauss and Student dependence structures. Figure 12 shows the $\mathcal{H}_S - A$ diagrams for observed and Gaussian resampled data and highlights that \mathcal{H}_S increases as A increases. The resampled data exhibit higher \mathcal{H}_S values, especially for moderately high quantiles (i.e., below or equal to $x_{0.99}$) and time scales below 24 h. The \mathcal{H}_S values confirm the difference between the rainfall in summer (JJA) and the other seasons. Similar results were found for the Student resampled data.

Since the practical information we are interested in is the rainfall amount, which impacts on the actual risk

Fig. 11 $\rho_{t_1} - \rho_{t_2}$ diagrams for gridded data at $0.25^\circ \times 0.25^\circ$ spatial resolution



assessment procedures, the loss of information related to Gaussian or Student spatial dependence structures was quantified by the mean and standard deviation of the rainfall amount summed over each triple. Figure 13 shows that the Gaussian dependence structure implies a systematic underestimation of the average rainfall amount on triples of locations, the bias reducing for the highest thresholds and coarse aggregation time scales, for which however the sample size reduces and the uncertainty increases. Note that Student dependence structure yields similar results, with a smaller bias for the highest quantile

thresholds ($x_{0.99}$ and $x_{0.995}$) and time scales below six hours (figures not shown). Both models reproduce rather well the standard deviation rainfall amount summed over each triple (figures not shown). Generally, we observed that the Student model (with tail dependence) produces a slight inflation of the average rainfall amount on triples, which improves the performance at fine time scales but gives a positive bias when the data are upscaled.

Daily and coarser time scales are explored using the central-eastern Europe data set. Figure 14 highlights a scaling relationship between \mathcal{H}_S and A with a break point

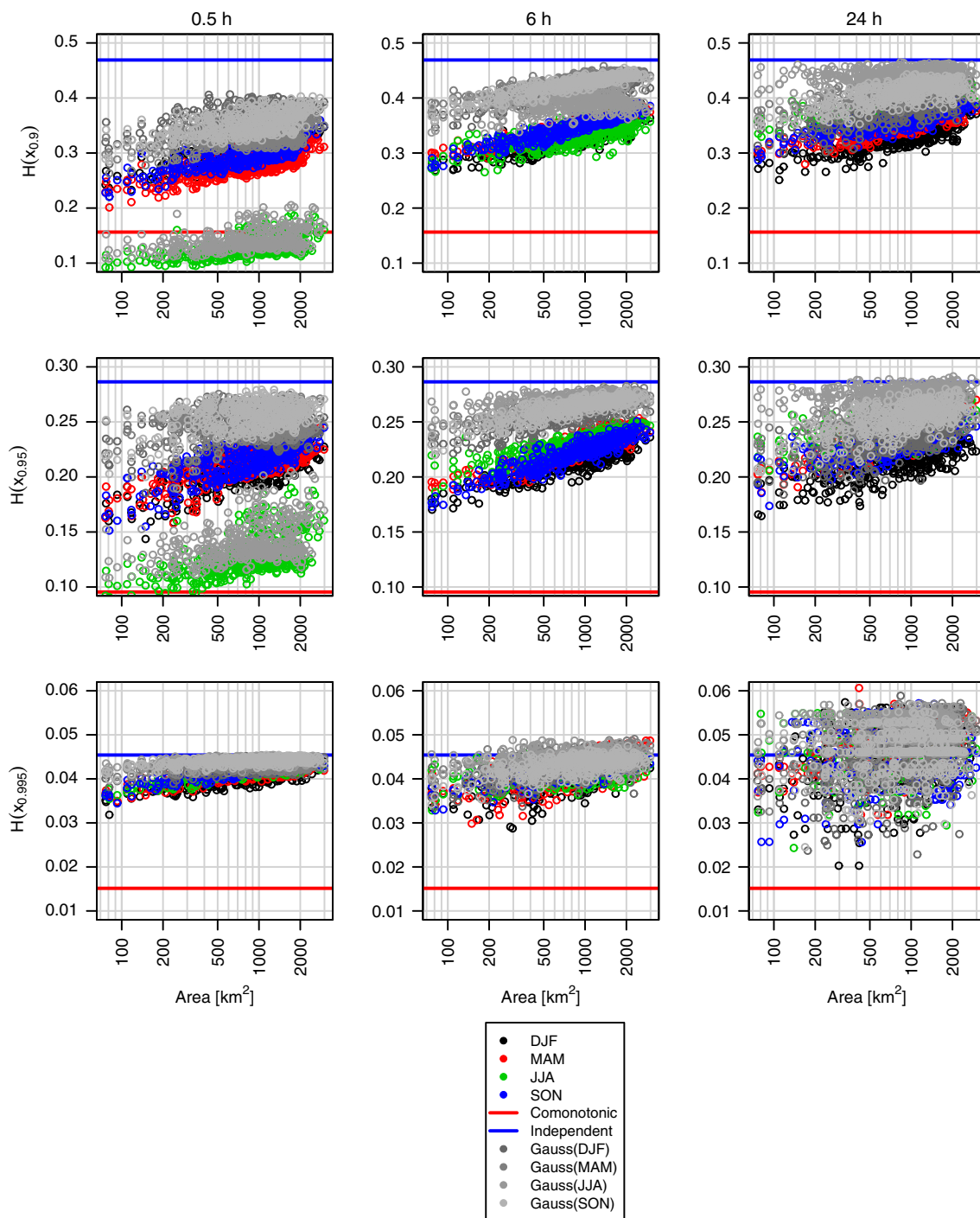


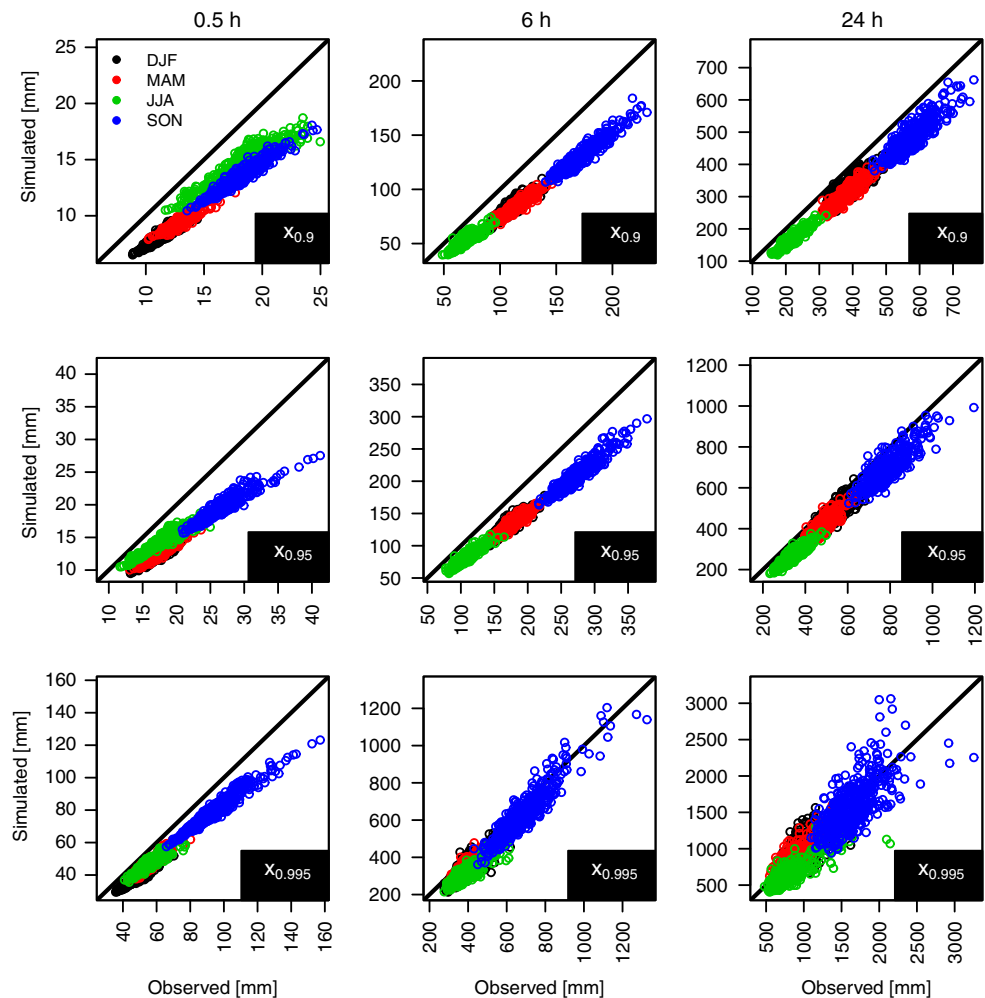
Fig. 12 Binary entropy on triples for Umbria rainfall data

around 20,000 km² corresponding to mutual inter-site distances of ≈280 km; beyond this distance the entropy is almost independent of the distance and approaches the limiting value corresponding to mutual independence. Figure 15 completes the picture showing that the negative bias observed at sub-daily time scales persists up to 2-day time scale, whereas it is smoothed out at coarser scales.

5 Discussion

When mathematical/statistical models and tools are introduced in applied disciplines, such as hydrology, theoretical limits and basic assumptions can be overlooked in the process of generating results. The upper tail dependence λ_U and its estimators seem to be some of these concepts. Since

Fig. 13 Average rainfall amount on triples for Umbria data. “Simulated” refers to resampled data with latent Gauss dependence structure obtained by the simulation algorithm described in Sect. 2.3



the simultaneous occurrence of extreme events has a strong impact in practical applications ranging from risk management to insurance strategies, the interest in empirical results concerning rainfall, flood and drought data (to mention just a few) has drawn attention away from the true asymptotic nature of λ_U and the reliability of the estimators applied to the generally short records of hydrological variables.

However, since UTD is an asymptotic property similar to other characteristics such as the univariate EV behavior and long range dependence, making inference on UTD requires large data sets and particular care. As for an example, Papalexiou and Koutsoyiannis (2013) and Serinaldi and Kilsby (2014b) highlighted the concealing effect of the (short) record length on the recognition of the nature of the univariate tails of rainfall extremes. The finite-sample properties (bias and uncertainty) of the statistical tools used in such analyses must always be checked, especially if they are borrowed from theoretical literature without specific preliminary checks accounting for the

requirements of hydrological studies and the properties of the data at hand. Moreover, often these tools are tested on large data sets (thousands) or high frequency econometric or biometric records (e.g., Poon et al. 2003; Wu et al. 2012; Cai et al. 2013). Therefore, blindly adopting the same technology for analyzing a few dozens of annual maxima or zero-inflated rainfall records might be expected to return misleading results.

In this respect, it should be noted that the set of diagnostics tested in this study is not exhaustive as other indices and tests for tail (in)dependence have been proposed in the literature (see e.g., Bacro and Toulemonde 2013 for a recent review). However, such alternative methods are generally affected by the same problems discussed in this study. For example, Coles et al. (1999) complemented λ_U with a complementary index $\bar{\lambda}_U$ related to the joint survival function and set up an inference procedure based on the pair $(\lambda_U, \bar{\lambda}_U)$ so that $(\lambda_U > 0, \bar{\lambda}_U = 1)$ indicates asymptotic dependence, and the value of λ_U determines a measure of strength of dependence within the class of asymptotically

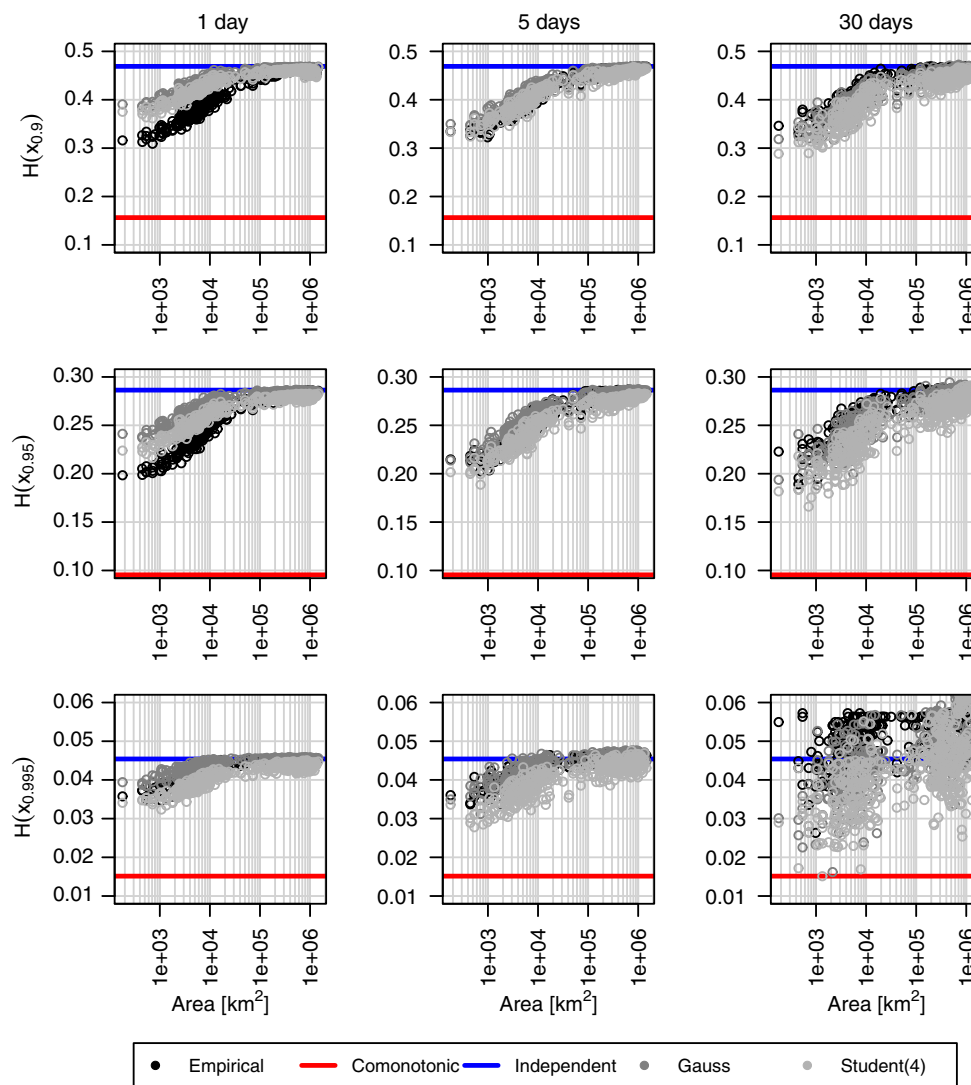


Fig. 14 Binary entropy on triples for central-eastern Europe rainfall data

dependent distributions; alternatively, $(\lambda_U = 0, \bar{\lambda}_U < 1)$ indicates asymptotic independence, and the value of $\bar{\lambda}_U$ determines the strength of dependence within the class asymptotically independent distributions. Unfortunately, this procedure still suffers the bias and uncertainty of the estimators of λ_U and $\bar{\lambda}_U$ as shown by Coles et al. (1999) for the Gauss dependence structure. Coles et al. (1999) also highlighted that the inference procedure on finite-size samples requires the definition of the confidence intervals of the estimates of λ_U and $\bar{\lambda}_U$, which are obtained by the delta method and rely on three assumptions: (1) independence of the observations, (2) each marginal distribution is estimated exactly by its empirical distribution function, and (3) the sampling distribution of a proportion is well approximated by its asymptotic distribution (see also Poon et al. 2003). Since these properties are rarely fulfilled (if ever), resulting in general underestimation of the width of

the confidence intervals, Coles et al. (1999) explicitly warned about the reliability of conclusions drawn from these intervals. The analyses of real-world environmental data discussed by Coles et al. (1999) also confirm the uncertainty affecting the results and the difficulty to distinguish between asymptotic dependence and asymptotic independence.

As far as tests of independence are concerned, Bacro and Toulemonde (2013) suggested to distinguish between approaches related to a sample from the bivariate EV distribution and approaches which can be used on distributions in a domain of attraction of an EV distribution. The first class comprises the tests proposed by Falk and Michel (2006), Bacro et al. (2010) and Ramos and Ledford (2005), whereas the second class includes the tests developed by Zhang (2008) and Hüsler and Li (2009), for instance. Even though we tested only some of the tests in both classes, our

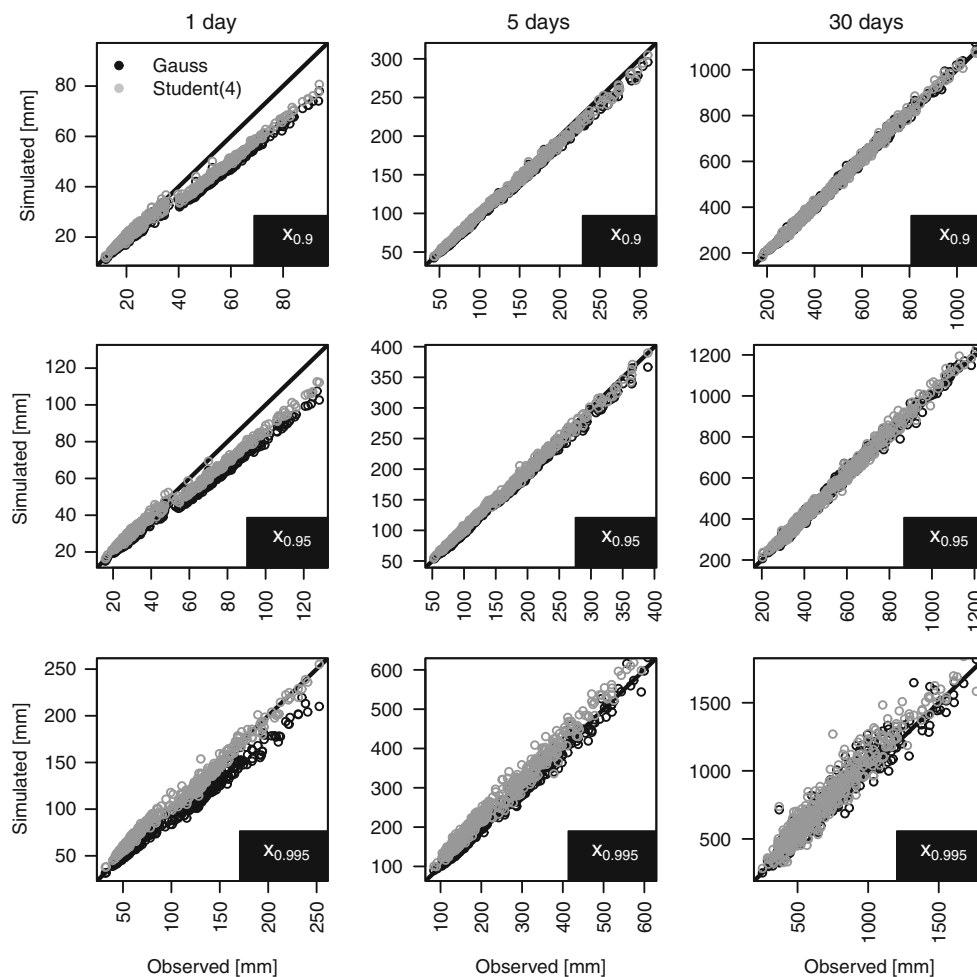


Fig. 15 Average rainfall amount on triples for central-eastern Europe data

conclusions about the effect of the sampling uncertainty and model misspecification hold true also for other tests. For example Bacro et al. (2010) recognized the slow convergence (and consequent lack of power) of their madogram test in the case of the Gauss model as the overall correlation increases (>0.5), which is coherent with our results concerning the difficulty of discriminating between limit dependence and independence for joint distribution with Gauss-like tail shape. The use of relatively large samples (hundreds or thousands of observations) for simulation-based power studies and real-world (usually econometric) examples (e.g., Poon et al. 2003; Hüsler and Li 2009; Bacro et al. 2010) also confirms that the minimal sample size needed to obtain reliable results is usually much larger than that of hydrological records of interest such as annual maxima or slightly longer peak-over-threshold data. Moreover, unlike well-behaved data such as time series of (filtered) returns (e.g., Bacro et al. 2010), hydrological data are affected by several factors (e.g., zeros' inflation, discretization due to finite resolution of

measurement devices, etc.) that can further conceal the true tail behavior.

6 Conclusions

In this study we have thoroughly revised the finite-sample properties of the most popular λ_U estimators and the recommendations and caveats about their application reported in the theoretical literature. Following these suggestions we have also studied the finite-sample performance of a suitable set of tests for tail independence under alternative models, as well as new alternative diagnostics. Finally, an extensive analysis of three rainfall data sets covering a wide range of spatio-temporal scales allowed us to draw general conclusions on the dependence properties of simultaneous extremes. Therefore, we obtained both methodological results which are general and valid for whatever type of data, and empirical results concerning the UTD behavior of rainfall fields, these being also rather

general in light of the spatio-temporal coverage of the data sets. The methodological results can be summarized as follows:

- (1) The λ_U^{SS} and λ_U^{CFG} estimators are generally biased and yield λ_U values strongly related to the overall correlation even if the underlying (true) dependence structure has zero tail dependence. These estimators are not only biased toward positive tail dependence but also affected by high uncertainty even for samples with large size (compared with the typical record length of hydrological data). In particular, it should be noted that λ_U^{CFG} is actually a parametric estimator as it relies on the assumption that the underlying structure of dependence is EV. Our Monte Carlo experiments highlight that this assumption is not weak and the estimator systematically points to EV models, being almost insensitive to the true nature of the tail dependence structure. Therefore, this estimator must be used with great care and possibly avoided, when the true dependence structure is unknown. Since its application is widespread because of its closed form expression (which does not need iterative algorithms), and several results reported in the hydrological literature rely on it, we argue that those results are questionable and must be carefully reconsidered.
- (2) Since tests for tail (in)dependence are absolutely mandatory for every λ_U estimation (Schmidt 2003; Frahm et al. 2005; Schmidt and Stadtmüller 2006), we have provided an extensive power study of five formal statistical tests proposed in the literature. We have checked the performance of these tests for dependence structures with zero and positive theoretical UTD and the whole range of possible positive correlation values using both a relatively large sample size (i.e. 1,000) and a typical hydrological record length (75; mimicking a typical set of annual maxima). Simulation results showed that Fisher test exhibits a poor performance and must be discarded, whereas Neyman–Pearson, Kolmogorov–Smirnov, χ^2 , and Zhang tests perform satisfactorily. In more detail, Kolmogorov–Smirnov and χ^2 are almost redundant as they use two different criteria to measure the discrepancy between the empirical and theoretical distributions of the same test statistic. On the other hand, Neyman–Pearson and Zhang tests rely on different rationales. Since the performance depends on the value of the overall correlation and sample size, we do not provide a ranking of these tests, but recommend the use of Neyman–Pearson, Zhang and at least one of Kolmogorov–Smirnov and

χ^2 in order to provide a cross-check based on different criteria.

- (3) Binary correlations and binary entropy on triples provide alternative diagnostics based on finite samples. Unlike λ_U estimators, binary correlation was shown to be almost unbiased but is unavoidably affected by large uncertainty when the record length has the typical values of hydrological data sets. On the other hand, the binary entropy on triples is an attempt to measure the so-called high-order dependence moving from pairwise mutual relationships to higher dimensional relationships which can reflect or be responsible for more complex interactions characterizing the dynamics of hydrological phenomena such as storms and floods. We anticipate that binary entropy and λ_U are formally related; however, these aspects are still under study and further results will be communicated in the future.

As far as the empirical results corresponding to the joint upper tail behavior of rainfall fields are concerned, we can conclude:

- (1) Looking at pairwise UTD values, both λ_U versus inter-site distances and $\tau_K - \lambda_U$ diagrams exhibit the typical patterns already reported in the literature, and confirm the strong relationship between λ_U estimates and the overall pairwise correlation values as well as the bias of the estimators. These results hold true for every data set across a broad range of spatio-temporal scales from 30 min and 1 km to 30 days and ≈ 3000 km.
- (2) The formal statistical tests for upper tail independence confirm the above results pointing to the existence of possible upper tail dependence in a number of cases much smaller than the expected. These cases correspond to nearby records characterized by strong overall correlation, which affect the tests' outcomes similarly to λ_U estimators.
- (3) The binary correlation computed over a suitable set of quantile thresholds further confirms the lack of empirical evidence for a general presence of the upper tail dependence in rainfall fields. In more detail, the empirical $\rho_{i1} - \rho_{i2}$ patterns are closer to the Gauss theoretical curves than EV models. Given the substantial lack of bias of the binary correlation estimators, this result is deemed reliable (keeping in mind the uncertainty of the estimates). Moreover, it should be noted that the small difference between the theoretical curves corresponding to Gauss and EV models for the smallest quantile thresholds does not allow us to recognize the true behavior of the upper tails. Therefore, $\rho_{i1} - \rho_{i2}$ must be applied to a set of

thresholds in order to highlight the evolution of the correlation in the upper tail region of the joint distribution and obtain insight about UTD.

- (4) The longer the aggregation time the closer the empirical entropy comes to the values corresponding to the latent Gauss or Student dependence structures. However, the entropy corresponding to these models is higher than the observed, indicating that there must be a stronger organization of extreme rainfall which is not captured by any of the previous models. Even though the introduction of tail dependence by the Student copula reduces the entropy (as expected) this is not enough to explain the above mentioned organization. On the other hand, in light of the results of the UTD analysis, there is no empirical evidence to justify the use of tail dependent models, which therefore assume the role of a fix rather than an explanation.
- (5) The analysis of the rainfall accumulation over triples of locations further confirms the loss of information resulting from the use of Gauss and Student models, thus meaning that UTD is not enough to explain the spatial structure of rainfall extremes. The underestimation of accumulated rainfall is more evident at sub-daily time scales, becoming negligible at coarser time scales. Indeed, a sort of convergence toward meta-elliptical dependence structures (such as Gauss and Student) might be expected because of the averaging and smoothing effect of the upscaling process.

Acknowledgments This work was supported by the Willis Research Network and Engineering and Physical Sciences Research Council (EPSRC) “UK Infrastructure Transitions Research Consortium” Grant EP/I01344X/1. The authors wish to thank Prof. Zhengjun Zhang (University of Wisconsin, US) for his advice about the quotient correlation test, and Prof. Qiang Zhang (Sun Yat-sen University, China) and three anonymous reviewers for their remarks and criticisms that helped improve the quality of the original manuscript. The analyses were performed in R (R Core Team 2013) by using the contributed packages POT (Ribatet 2006), mvtnorm (Genz and Bretz 2009; Genz et al. 2011), and copula (Yan 2007; Kojadinovic and Yan 2010). The authors and maintainers of this software are gratefully acknowledged.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- AghaKouchak A, Ciach G, Habib E (2010) Estimation of tail dependence coefficient in rainfall accumulation fields. *Adv Water Resour* 33(9):1142–1149
- AghaKouchak A, Sellars S, Sorooshian S (2013) Methods of tail dependence estimation. In: AghaKouchak A, Easterling D, Hsu K, Schubert S, Sorooshian S (eds) *Extremes in a changing climate, water science and technology library*, vol 65. Springer, Dordrecht, pp 163–179
- Bacro J, Toulemonde G (2013) Measuring and modelling multivariate and spatial dependence of extremes. *J Soc Fr Stat* 154(2):139–155
- Bacro J, Bel L, Lantuéjoul C (2010) Testing the independence of maxima: from bivariate vectors to spatial extreme fields. *Extremes* 13(2):155–175
- Bárdossy A (2006) Copula-based geostatistical models for groundwater quality parameters. *Water Resour Res* 42(11):WR004,754
- Bárdossy A, Li J (2008) Geostatistical interpolation using copulas. *Water Resour Res* 44(7):WR006,115
- Bárdossy A, Pegram G (2009) Copula based multisite model for daily precipitation simulation. *Hydrol Earth Syst Sci* 13(12):2299–2314
- Bárdossy A, Pegram G (2013) Interpolation of precipitation under topographic influence at different time scales. *Water Resour Res* 49(8):4545–4565
- Cai JJ, Fougères A, Mercadier C (2013) Environmental data: multivariate Extreme Value Theory in practice. *J Soc Fr Stat* 154(2):178–199
- Capéraà P, Fougères AL, Genest C (1997) A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* 84(3):567–577
- Coles S, Heffernan J, Tawn J (1999) Dependence measures for extreme value analyses. *Extremes* 2(4):339–365
- Falk M, Michel R (2006) Testing for tail independence in extreme value models. *Ann Inst Stat Math* 58(2):261–290
- Falk M, Reiss RD (2005) On Pickands coordinates in arbitrary dimensions. *J Multivar Anal* 92(2):426–453
- Frahm G, Junker M, Schmidt R (2005) Estimating the tail-dependence coefficient: properties and pitfalls. *Insur Math Econ* 37(1):80–100
- Genz A, Bretz F (2009) Computation of multivariate normal and t probabilities. *Lecture Notes in Statistics*. Springer, Heidelberg
- Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T (2011) mvtnorm: multivariate normal and t distributions. R package version 0.9-9991. <http://CRAN.R-project.org/package=mvtnorm>
- Ghizzoni T, Roth G, Rudari R (2010) Multivariate skew-t approach to the design of accumulation risk scenarios for the flooding hazard. *Adv Water Resour* 33(10):1243–1255
- Haylock MR, Hofstra N, Klein Tank AMG, Klok EJ, Jones PD, New M (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *J Geophys Res* 113(D20):D20,119
- Hüsler J, Li D (2009) Testing asymptotic independence in bivariate extremes. *J Stat Plan Inference* 139(3):990–998
- Janga Reddy M, Singh VP (2014) Multivariate modeling of droughts using copulas and meta-heuristic methods. *Stoch Environ Res Risk Assess* 28(3):475–489
- Joe H, Smith RL, Weissman I (1992) Bivariate threshold methods for extremes. *J R Stat Soc Ser B (Methodol)* 54(1):171–183
- Klein Tank AMG, Wijngaard JB, Können GP, Böhm R, Demarée G, Gocheva A, Mileta M, Pashiardis S, Hejkrlik L, Kern-Hansen C, Heino R, Bessemoulin P, Müller-Westermeier G, Tzanakou M, Szalai S, Pálsdóttir T, Fitzgerald D, Rubin S, Capaldo M, Maugeri M, Leitass A, Bukantis A, Aberfeld R, van Engelen AFV, Forland E, Miletus M, Coelho F, Mares C, Razuvaev V, Nieplova E, Cegnar T, Dahlström B, Moberg A, Kirchhofer W, Ceylan A, Pachaliuk O, Alexander LV, Petrovic P (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *Int J Climatol* 22(12):1441–1453

- Kojadinovic I, Yan J (2010) Modeling multivariate distributions with continuous margins using the copula R package. *J Stat Softw* 34(9):1–20
- Nelsen RB (2006) An introduction to copulas, 2nd edn. Springer, New York
- Papalexiou SM, Koutsoyiannis D (2013) Battle of extreme value distributions: a global survey on extreme daily rainfall. *Water Resour Res* 49(1):187–201
- Pickands J III (1981) Multivariate extreme value distributions. In: Proceedings of the 43th session of the International Statistical Institute, Buenos Aires
- Poon S, Rockinger M, Tawn J (2003) Modelling extreme-value dependence in international stock markets. *Stat Sin* 13:929–953
- Poulin A, Huard D, Favre AC, Pugin S (2007) Importance of tail dependence in bivariate frequency analysis. *J Hydrol Eng* 12(4):394–403
- R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org/>, ISBN 3-900051-07-0
- Ramos A, Ledford A (2005) Regular score tests of independence in multivariate extreme values. *Extremes* 8(1–2):5–26
- Requena AI, Mediero L, Garrote L (2013) A bivariate return period based on copulas for hydrologic dam design: accounting for reservoir routing in risk estimation. *Hydrol Earth Syst Sci* 17(8):3023–3038
- Ribatet MA (2006) A user's guide to the POT package (Version 1.0). <http://cran.r-project.org/>
- Schmidt R (2003) Dependencies of extreme events in finance. PhD thesis, Fakultät für Mathematik und Wirtschaftswissenschaften, Universität Ulm, Ulm (DE)
- Schmidt R, Stadtmüller U (2006) Non-parametric estimation of tail dependence. *Scand J Stat* 33(2):307–335
- Serinaldi F (2008) Analysis of inter-gauge dependence by Kendall's τ_K , upper tail dependence coefficient, and 2-copulas with application to rainfall fields. *Stoch Environ Res Risk Assess* 22(6):671–688
- Serinaldi F (2013) An uncertain journey around the tails of multivariate hydrological distributions. *Water Resour Res* 49(10):6527–6547
- Serinaldi F (2014) Dismissing return periods! *Stoch Environ Res Risk Assess*. doi:[10.1007/s00477-014-0916-1](https://doi.org/10.1007/s00477-014-0916-1)
- Serinaldi F, Kilsby CG (2013) The intrinsic dependence structure of peak, volume, duration, and average intensity of hyetographs and hydrographs. *Water Resour Res* 49(6):3423–3442
- Serinaldi F, Kilsby CG (2014a) Simulating daily rainfall fields over large areas for collective risk estimation. *J Hydrol* 512:285–302
- Serinaldi F, Kilsby CG (2014b) Rainfall extremes: toward reconciliation after the battle of distributions. *Water Resour Res* 50(1):336–352
- Serinaldi F, Bonaccorso B, Cancelliere A, Grimaldi S (2009) Probabilistic characterization of drought properties through copulas. *Phys Chem Earth Parts A/B/C* 34(10–12):596–605
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–423
- Villarini G, Serinaldi F, Krajewski WF (2008) Modeling radar-rainfall estimation uncertainties using parametric and non-parametric approaches. *Adv Water Resour* 31(12):1674–1686
- Volpi E, Fiori A (2012) Design event selection in bivariate hydrological frequency analysis. *Hydrol Sci J* 57(8):1506–1515
- Wu J, Zhang Z, Zhao Y (2012) Study of the tail dependence structure in global financial markets using extreme value theory. *J Rev Glob Econ* 1:62–81
- Yan J (2007) Enjoy the joy of copulas: with a package copula. *J Stat Softw* 21(4):1–21
- Zhang Z (2008) Quotient correlation: a sample based alternative to Pearson's correlation. *Ann Stat* 36(2):1007–1030